ORFeome phage display to identify microbiome-derived biomarker candidates in the context of inflammatory disorders of the gut

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Jonas Max Zantow
aus Filderstadt

1. Referent: apl. Professor Dr. Michael Hust
2. Referent: Professor Dr. Stefan Dübel
3. Referentin: Professorin Dr. Susanne Engelmann
eingereicht am: 13.07.2016
mündliche Prüfung (Disputation) am: 24.10.2016

Druckjahr 2016

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Zantow, J., Just, S., Lagkouvardos, I., Kisling, S., Dübel, S., Lepage, P., Clavel, T. and Hust, M. (2016). Mining gut microbiome oligopeptides by functional metaproteome display. *Sci Rep* **6**, 34337

Zantow, J., Moreira, G.M.S.G., Dübel, S., Hust, M. (2016). ORFeome phage display. (eingereicht)

Tagungsbeiträge

Zantow, J. and Hust, M.: A vaccine pipeline - phage display for biomarker identification and generation of human antibodies for diagnostics and therapy. (Vortrag). BIT's 8th Annual World Congress of Vaccine, Dalian, China (2016).

Zantow, J., Just, S., Sokol, H., Haller, D., Lepage, P, Clavel, T. and Hust, M.: Functional Metaproteome Display - Applying phage display to discover novel serologic biomarkers for Inflammtory Bowel Disease. (Vortrag und Poster).
4th Inflammatory and Immunological Biomarkers Summit, San Diego, USA (2015).

Zantow, J., Just, S., Sokol, H., Haller, D., Lepage, P., Clavel, T. and Hust, M.: Metaproteome Display: A novel approach for biomarker discovery in Inflammatory Bowel Disease. (Vortrag). 2nd World Congress on Targeting Microbiota, Paris, Frankreich (2014).

Table of Contents

Abbreviations

%	Percent
+	Wildtype, positive
°C	Degree celsius
µg	Microgram(s)
µL	Microliter(s)
µM	Micromolar
ACCA	Anti-chitobioside antibodies
ALCA	Anti-laminaribioside antibodies
AMCA	Anti-mannobioside antibodies
ANCA	Anti-neutrophil cytoplasmic antibodies
APS	Ammonium persulfate
ARE	AU-rich elements
ASCA	Anti-Saccharomyces cerevisiae antibodies
AUC	Area under curve
BLAST	Basic Local Alignment Search Tool
bp	Basepair(s)
BSA	Bovine serum albumine
CD	Crohn's disease
cDNA	Complementary DNA
cfu	Colony forming unit(s)
conc.	Concentration
conv.	Conventional
CRP	C-reactive protein
DAB	3,3′-Diaminobenzidine
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetra-acetic acid
ELISA	Enzyme-linked immunosorbent assay
ESR	Erythrocyte sedimentation rate
FACS	Fluorescence-activated cell sorting
FMT	Fecal microbiota transplantation
Fwd.	Forward
g	Gram(s), earth acceleration
Gbp	Gigabasepair(s)
gDNA	Genomic DNA
gIII	M13 minor coat protein III gene
GM-CSF	Granulocyte macrophage colony-stimulating factor
GP2	Glycoprotein 2
h	Hour(s)
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IBD	Inflammatory bowel disease

IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranosid
kb	Kilobase(s)
kDa	Kilodalton(s)
kV	Kilovolt(s)
M	Molar
mA	Milliampere(s)
MDA	Multiple displacement amplification
min	Minute(s)
mL	Milliliter(s)
mM	Millimolar
mRNA	Messenger RNA
MS	Mass spectrometry
MTP	Microtiter plate
MW	Molecular weight
MWCO	Molecular weight cut-off
MyD88	Myeloid differentiation primary response protein 88
ng	Nanogram(s)
NGS	Next generation sequencing
nm	Nanometer(s)
nt	Nucleotide(s)
OD_{600}	Optical density at wave length 600 nm
OmpC	Outer membrane protein C
OmpW	TonB-linked outer membrane protein
ORF	Open reading frame
PAB	Anti-pancreas antibodies
PAGE	Polyacrylamide gelelectrophoresis
pANCA	Atypical perinuclear antineutrophilic cytoplasmic antibodies
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pelB	Pectate lyase B leader sequence
pIII	M13 minor coat protein III
pVIII	M13 major coat protein VIII
Rev.	Reverse
RNA	Ribonucleic acid
Rpm	Rounds per minute
rRNA	Ribosomal RNA
s	Second(s)
SAICAR	Phosphoribosylamino-imidazolesuccino-carboxamide
SDS	Sodium dodecyl sulfate
SEREX	Serologic expression cloning
SPF	Specific pathogen free
TEMED	N,N,N',N'-Tetramethylethylendiamin

tetR	Tetracycline repressor
TNF	Tumor necrosis factor
U	Unit(s)
UC	Ulcerative colitis
UV	Ultra-violet
V	Volt(s)
v/v	Volume per volume
w/v	Weight per volume
wt	Wildtype
Δ	Delta, deletion
λ	Phage lambda, wave length
3' UTR	3' untranslated region

1 Introduction

1.1 Biomarkers

Biomarkers are indispensable tools for diagnostics and clinical disease management and can be measured in different body fluids. Due to the ease of collection, blood serum is the most convenient source for biomarker measurements, it circulates through all body regions and tissues and contains numerous proteins that can be measured for diagnostic purposes. Immunoglobulins in particular comprise a group of serum proteins which can be indicative of specific diseases. Countless diagnostic ELISAs and other assays rely on the detection of specific immunoglobulins. Technological advances such as protein microarrays, peptide microarrays or Luminex multiplex assays today also allow fast and cost-efficient measurement of complex patterns of antibody responses (I. H. Khan et al., 2011; Ravindran et al., 2014; Robinson et al., 2002; Schirwitz et al., 2012; van den Berg et al., 2011). However, the identification of the antigens that trigger these specific systemic antibody responses is still difficult and limits the application of these technologies.

1.2 Biomarker discovery tools and their limitations

In order to develop efficient diagnostics and identify target structures of specific antibody responses indicative for disease or inflammatory disorders, several technologies have been developed for biomarker discovery but are limited in one or the other way.

1.2.1 Classical proteomics

Classical proteomic studies to identify pathogen proteins for diagnostic applications typically rely on the cultivation of the pathogen followed by 2D-PAGE and immunoblot of the proteome using patient sera and protein identification by mass spectrometry (MS) (Delvecchio et al., 2006; Jacobsen, Meens, Baltes, & Gerlach, 2005; LaFrentz, LaPatra, Call, Wiens, & Cain, 2011). But proteome analysis after *in vitro* cultivation does not allow the identification of proteins that are only expressed during host-pathogen interactions. To identify immunogenic proteins that are specifically important in pathogenesis, the pathogen has to be cultivated in direct contact with the host to ensure gene expression patterns reflecting the true pathogenicity situation. While this approach has been applied with some success, its inherent problem is that the pathogen proteome will most likely be overwhelmed by that of the host (Zhang, Chromy, & McCutchen-Maloney, 2005). Further,

two other groups of potential biomarkers also usually fail to be identified by classical proteome analysis. Proteins with a molecular mass smaller than 10 kDa or weakly expressed proteins may not be identified by 2D-PAGE/MS (Beranova-Giorgianni, 2003; Rabilloud, Chevallet, Luche, & Lelong, 2010; Urquhart, Cordwell, & Humphery-Smith, 1998).

1.2.2 Protein and peptide arrays

As an alternative, arrays with several hundred proteins spotted on glass slides and subsequently incubated with patient sera have been used for biomarker discovery. Immunogenic proteins from pathogens (Danckert, Hoppe, Bier, & von Nickisch-Rosenegk, 2014; Hoppe, Bier, & von Nickisch-Rosenegk, 2013; Kunnath-Velayudhan et al., 2010), *E. coli*-derived biomarkers for inflammatory bowel disease (IBD) (C. S. Chen et al., 2009), and exposure related antibody responses to proteins from *Plasmodium falciparum* (Helb et al., 2015) were identified by protein arrays. Even though microarrays allow the screening of several hundred target proteins at a time, the technology is laborious and expensive as each protein must be produced in a recombinant manner. Peptide microarrays can be produced with potentially millions of individual sequences (Schirwitz et al., 2012) allowing to cover entire proteomes, but they are still limited to only detect antibodies binding to short linear epitopes.

1.2.3 Serologic expression cloning

Gene library based screening technologies potentially offer a workaround for the above limitations. Serologic expression cloning (SEREX), a technology developed in the 1990s, relies on the expression of cDNA libraries in *E. coli* using lytic phage vectors (Y. T. Chen, Gure, & Scanlan, 2005). The cells are then plated and produce cDNA library encoded proteins intracellularly. The λ phage causes cell lysis and the resulting plaques containing the expressed proteins are transferred to a membrane that is stained using patient sera. The SEREX technology was initially developed for serologic analysis of tumor cDNA expression libraries (Türeci, Sahin, & Pfreundschuh, 1997), but was also used to identify bacterial flagellin as a dominant antigen associated with Crohn's disease using an expression library constructed from metagenomic DNA of the caecum of a transgenic mouse model for experimental colitis (Lodes et al., 2004). However, handling of lytic

phage, maintenance of the phage library, and laborious screening procedures are major drawbacks of this technology.

1.2.4 *E. coli* display

Instead of cytoplasmic expression of recombinant proteins and release by cell lysis, large libraries may also be screened by displaying peptides or proteins on the surface of selectable particles. Random peptide libraries displayed on *E. coli* cells were enriched for displayed epitopes specifically reactive with sera from celiac disease patients using fluorescent activating cell sorting (FACS). Subsequent directed peptide evolution and further sorting steps revealed biomarker candidates for the disease (Ballew et al., 2013). However, working with sera and *E. coli* cells as display chassis requires additional subtraction steps using control immunoglobulins, since antibodies reactive towards *E. coli* proteins usually are present. This drawback may be especially problematic for the identification of bacterial antigens in the context of gut-related diseases.

1.3 Phage display

The biomarker discovery approach of the present study depends on phage display, a technology initially invented by G.P. Smith in 1985 (Smith, 1985). To allow the display of heterologous proteins on the surface of phage particles, they are fused to a structural M13 phage protein whereas the corresponding coding DNA sequence of the heterologous protein is packaged in the same phage particle.

Phage display can be used to identify protein interaction partners based on the binding characteristics of the presented heterologous protein. Therefore, the desired interaction partner is immobilized on a solid surface and subsequently incubated with the protein displaying phage particles. Stringent washing removes non-bound phage particles and the phage displaying proteins that specifically bind to the immobilized interaction partners are enriched at the solid surface. Elution of bound phage particles allows infection of *E. coli* to provide the coding DNA of the selected proteins for DNA sequencing.

The main application of phage display is the selection of human antibodies from antibody gene libraries (Breitling, Dübel, Seehaus, Klewinghaus, & Little, 1991; Hoogenboom, 2005; McCafferty, Griffiths, Winter, & Chiswell, 1990) and was used to generate antibodies for diagnostics (Fuchs et al., 2014), therapy (Mazumdar, 2009) as well as detection

reagents for basic research (Bradbury, Sidhu, Dübel, & McCafferty, 2011; Dübel, Stoevesandt, Taussig, & Hust, 2010).

However, phage display has also been used for epitope mapping of antibodies (Blüthner, Bautz, & Bautz, 1996; Zhou et al., 2013). Here, the antigen cDNA is randomly fragmented and cloned into a phage display vector. Based on the protein interaction between the antibodies and the displayed peptides, the DNA fragments coding for the epitope can be selected.

The phage display technology was further used for the identification of allergens (Crameri & Walter, 1999; Kodzius et al., 2003; Rhyner et al., 2004) and the identification of immunogenic proteins of pathogens, e.g. *Mycobacterium tuberculosis* (Liu et al., 2011), *Mycoplasma mycoides* (Miltiadou, Mather, Vilei, & Du Plessis, 2009), the rickettsia *Cowdria ruminantium* (Fehrsen & du Plessis, 1999) or the eukaryotic pathogen *Taenia solium* (Gonzalez et al., 2010) by cloning cDNA or whole genome fragments. However, non-directional cloning of cDNA or whole genome fragments into phage display vectors result in the majority of sequences not encoding any protein.

1.3.1 Enrichment of open reading frames in phage display libraries

Only one out of eighteen randomly cloned DNA fragments statistically result in correct open reading frames (ORFs). This means, only fragments cloned in the correct orientation, correct reading frame and fragments with the correct length do not lead to frameshifts. Additional stop codon containing “junk” sequences originate from non-coding DNA like gene regulatory elements. The “junk” content in libraries derived from eukaryotic genomic DNA is even more increased as introns account for a large portion of the eukaryotic genome. The propagation of phage without inserts or out-of-frame insert is more favored and in consequence affecting phage library composition and complexity as well as the proportion of protein displaying particles which dramatically hampers the panning success (Hust et al., 2006; Stratmann & Kang, 2005). Therefore, decreasing the portion of DNA fragments in the library that do not code for a correct protein sequence substantially increases the chance to eventually identify specific protein interactions.

To address this problem, the DNA libraries should be enriched for ORFs. Cloning cDNA upstream of a selection marker (e.g. β-lactamase conferring ampicillin resistance) in an

E. coli vector allows the enrichment of ORFs as only DNA fragments without stop codons and cloned in-frame with the β-lactamase gene allow the expression of β-lactamase and the corresponding *E. coli* clone can grow in selection medium supplemented with ampicillin (Seehaus, Breitling, Dübel, Klewinghaus, & Little, 1992). However, this method is tedious, too, since the inserts have to be subcloned into a phage display vector after ORF enrichment to allow selection of protein interaction partners (Faix et al., 2004).

For phage display, the protein of interest is displayed by fusion to an M13 phage coat protein, typically minor coat protein 3 (pIII). A phage particle has about five pIII copies and this protein provides the initial binding of the phage to the *E. coli* F pilus to initiate infection. Consequently, a phage particle without pIII is non-infectious and cannot be amplified. Due to their independent replication and regulation, phage display vectors are usually phagemids, i.e. plasmids that can be packaged into phage particles and contain the coding DNA for the foreign protein fused to the pIII gene but no other phage proteins. In order to complement the missing phage proteins like major coat protein pVIII and other necessary non-structural proteins, the *E. coli* cells have to be infected with a helperphage. Different antibiotics resistances and origins of replication (ori) allow the coexistence of phagemid and helperphage genome in the same *E. coli* cell and an attenuated packaging signal in the helperphage genome ensures that the phagemid is preferably packaged into the phage particle.

The enrichment for correct ORFs and subsequent functional selection using phage display without any further cloning steps was enabled by a special helperphage (Figure 1-1), referred to as “Hyperphage” (M13K07ΔgIII) (Rondot, Koch, Breitling, & Dübel, 2001; Soltes et al., 2007).

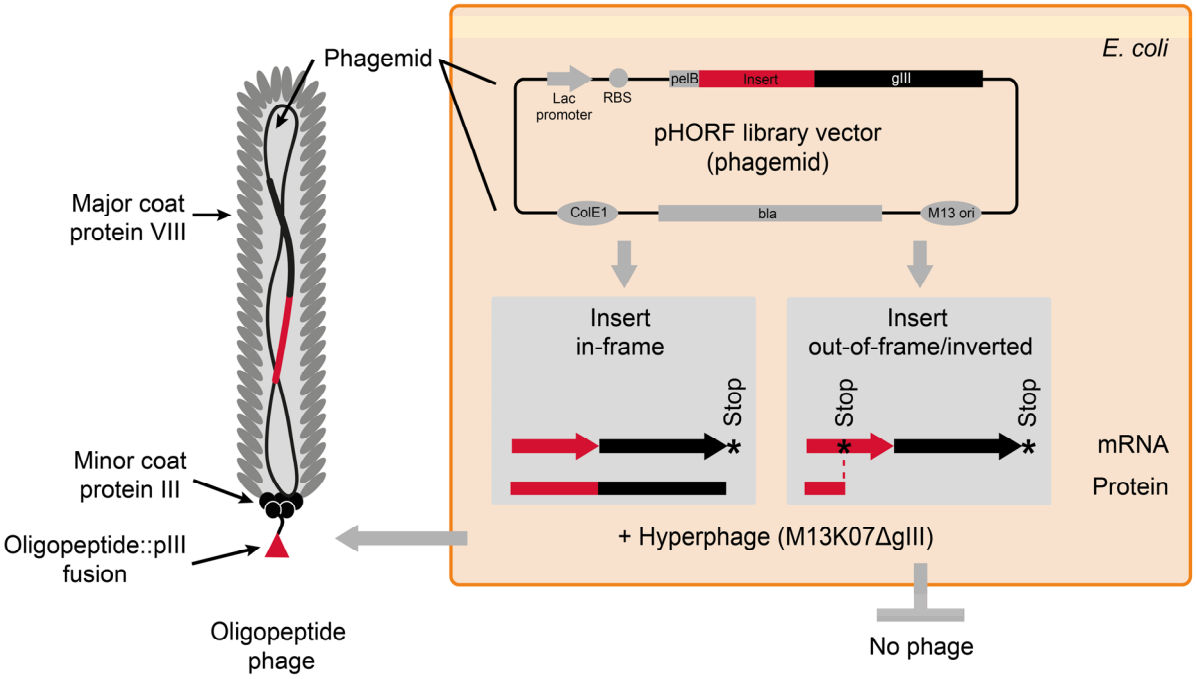


Figure 1-1: Schematic illustration of an oligopeptide phage, phagemid library vector and ORF enrichment using "Hyperphage". Phage display links genotype and phenotype, as the genetic information coding for the displayed protein or oligopeptide is contained on the phagemid packaged in the same phage particle. Non-directional cloning of DNA fragments only lead to a minor fraction of insert encoding proteins. When using a special helperphage ("Hyperphage", deleted pIII gene), the phagemid encoded pIII fusion protein is the only pIII source. Infectious phage particles are only assembled if the cloned DNA sequence is in-frame with the pIII gene and does not contain any stop codon leading to an enrichment of ORFs in a library. At the same time, the encoded protein is displayed on the phage particle as fusion protein with minor coat protein III (pIII).

"Hyperphage" is a unique helperphage as it has a deletion in the pIII gene (genotype pIII-) but carries functional pIII proteins and therefore is infectious (phenotype pIII+). Infection of *E. coli* provides it with all phage proteins except from pIII and consequently does not result in the production of infectious helperphage particles. Infection with "Hyperphage" allows the enrichment of ORFs encoded in suitable phagemids, as the phagemid is the only pIII source and is only produced if the randomly fragmented DNA cloned upstream of the pIII gene is in-frame with the pIII gene and without stop codon (Figure 1-1) ORF enriched phage display libraries (ORFeome phage display) can directly be used to select immunogenic proteins in a panning using patient sera (Figure 1-2). Specifically bound proteins or oligopeptides are usually enriched over 2-3 rounds of panning and immunogenic proteins as potential biomarker candidates are finally identified by monoclonal ELISA and DNA sequencing.

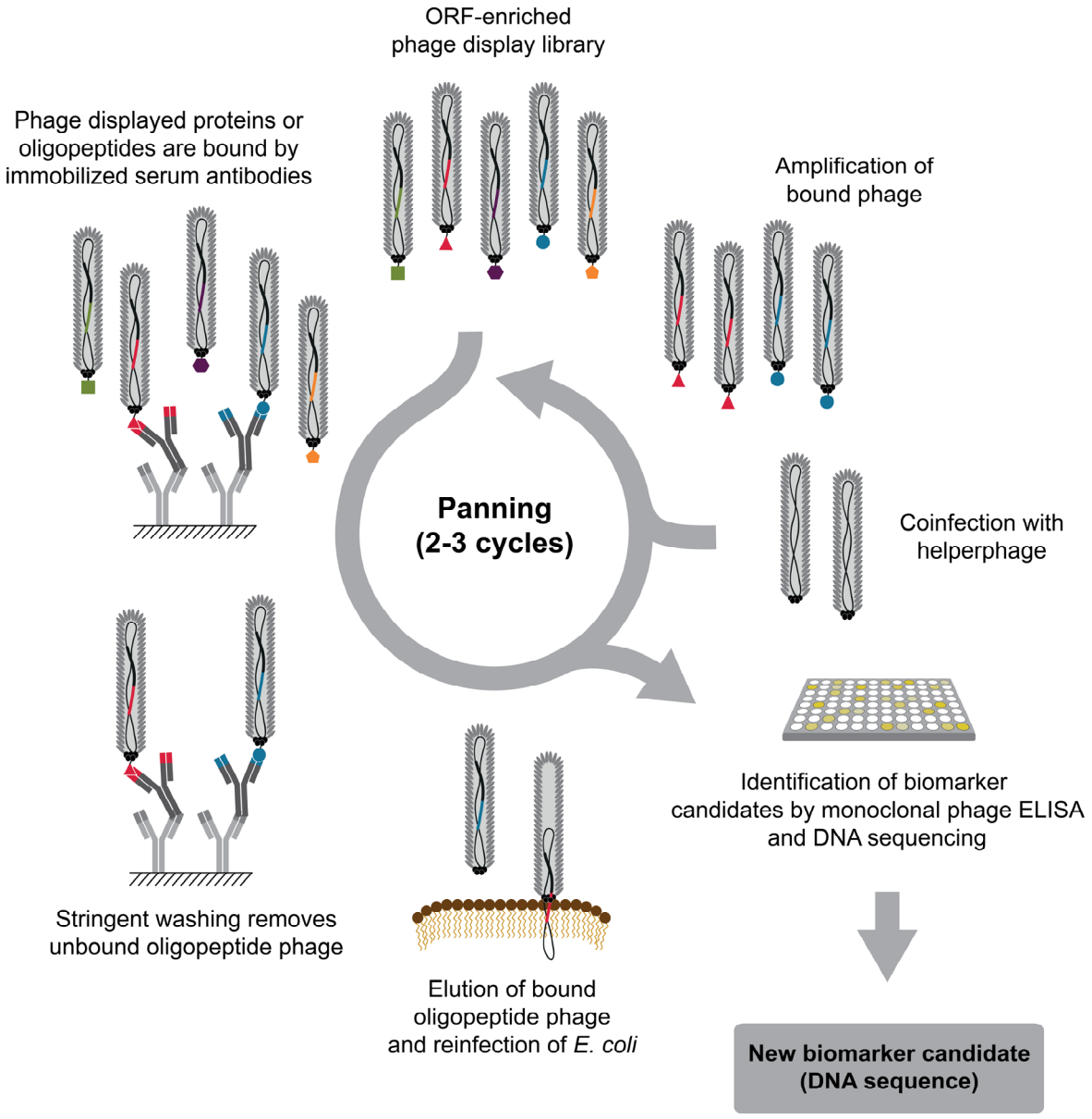


Figure 1-2: Schematic illustration of a panning procedure. ORF-enriched phage display library is applied to serum antibodies immobilized on a solid surface. Stringent washing removes non-bound oligopeptides. Elution and reinfection of *E. coli* allows the amplification of bound oligopeptide phage after coinfection with a helperphage. Monoclonal screening by ELISA identifies immune-reactive oligopeptides and the biomarker candidate is subsequently identified by DNA sequencing.

Thus, ORFeome phage display combines the positive selection of ORFs from whole genomes and metagenomes with the functional display of protein fragments coded by these sequences.

1.3.2 Applications of ORFeome phage display

The ORFeome phage display technology was used to identify novel immunogenic proteins from two different *Mycoplasma* species (Kügler et al., 2008; Naseem et al., 2010), *Salmonella* Typhimurium (Meyer et al., 2012) and *Neisseria gonorrhoeae* (Connor, Zantow, Hust, Bier, & von Nickisch-Rosenegk, 2016). The technology was also validated to work with eukaryotic cDNA, as immunogenic saliva proteins involved in tick feed were identified from *Ixodes scapularis* (Becker et al., 2015).

Despite ORFeome phage display can overcome certain limitations of other biomarker discovery technologies, proteins have to be able to be expressed in *E. coli* and displayed on phage particles. Further, ORFeome phage display usually only displays protein domains or oligopeptides and not whole proteins on the surface.

However, as ORFeome phage display is independent of cultivation and does not rely on sequenced or annotated genomes, it can be applied on complex microbial communities. In a recent study, a ruminal metasecretome was analyzed using phage display (Ciric et al., 2014). The authors cloned a rumen bacterial metagenome library in a phage display vector and enriched the library for secreted proteins by packaging with a pIII deficient helper phage (VCSM13d3, (Rakonjac, Jovanovic, & Model, 1997)). Only cloned gene fragments coding for endogenous signal peptides resulted in a pIII fusion protein and functional phage particles were sarkosyl resistant. The enriched metasecretome library was subsequently analyzed by next generation sequencing. The logical next step and aim of the present work was the use of metagenomic DNA for a functional metaproteome display applying the ORFeome phage display technology and selection of biomarker candidates based on protein-protein interactions.

1.4 Inflammatory bowel disease

Inflammatory bowel diseases (IBD) are idiopathic chronic relapsing inflammatory disorders of the gastro-intestinal tract characterized by bloody diarrhea and abdominal pain. The two major phenotypes of IBD are Crohn's disease (CD) and ulcerative colitis (UC). CD is characterized by a patchy transmural inflammation in ileum and colon whereas UC only affects the colon and rectum with continuous proximal moving mucosal inflammation.

IBD is a global burden with highest incidence in developed countries (Cosnes, Gower-Rousseau, Seksik, & Cortot, 2011; Molodecky et al., 2012). Estimated 2.5 - 3 million Europeans (Burisch, Jess, Martinato, & Lakatos, 2013), 233,000 Canadians (Rocchi et al., 2012) and 1.2 million individuals in the US (Kappelman, Moore, Allen, & Cook, 2013) are affected. The economic burden in 2014 was estimated to range between 14.6 and 31.6 billion USD in the US alone (Mehta, 2016).

Despite the etiology of IBD is not fully understood, it is believed to result from an aberrant immune response towards the gut microbiota in a genetically susceptible host. Genetic predisposition markers comprise mutations in genes associated with microbial sensing (Duerr et al., 2006; Hugot et al., 2001; Jostins et al., 2012). Moreover, shifts in microbiota composition are associated with IBD (Gevers et al., 2014; Manichanh et al., 2006), highlighting the crucial role of the microbiome in IBD pathogenesis.

The treatment of CD and UC is as individual-specific as the disease. Surgery in CD may become necessary to treat complications and colectomy can cure intestinal symptoms in UC but a cure without removing large parts of the intestines is not available. IBD management is mostly limited to treat the symptoms and to induce and maintain disease remission using chemical immunosuppressive and anti-inflammatory drugs (Bernstein, 2015; Triantafillidis, Merikas, & Georgopoulos, 2011) as well as biologicals such as the monoclonal anti-TNF antibodies infliximab, adalimumab and certolizumab pegol (Danese et al., 2014; Peyrin-Biroulet, 2010; Peyrin-Biroulet et al., 2008). Another monoclonal antibody, vendolizumab, was approved by FDA and EMA in 2014 for CD and UC treatment. Vendolizumab blocks α-integrins and prevents leukocytes to migrate into the intestinal mucosa (Raine, 2014). But biologicals are expensive and good biomarkers to identify drug responders and non-responders are needed.

With respect to the gut microbiome, antibiotics treatment was reported to have beneficial effects on disease activity in CD and UC (K. J. Khan et al., 2011). Addressing a pathogenic dysbiosis of the gut microbiome in IBD, the experimental treatment of fecal microbiota transplantation (FMT) has gained attention in the last years. A recent study reported FMT to induce remission in UC patients (Moayyedi et al., 2015) but more randomized trials are

needed to evaluate the efficacy of FMT for IBD treatment (Colman & Rubin, 2014; Moayyedi, 2016; Rossen et al., 2015).

1.4.1 Biomarkers in IBD

Biomarkers in IBD comprise different fecal markers such as lactoferrin and calprotectin as well as serologic markers like C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR). Besides CRP and ESR to assess an inflammatory state, specific humoral immune responses have been associated with IBD (Iskandar & Ciorba, 2012). These IgG and IgA responses comprise autoantibodies and a broad panel of antibodies against microbial antigens such as carbohydrate structures of yeast and bacterial proteins which were identified using different biomarker discovery approaches (Table 1-1).

The use of antibodies as diagnostic, prognostic or predictive biomarkers in IBD has extensively been reviewed (Bonneau et al., 2015; Ferrante et al., 2007; Peyrin-Biroulet, Standaert-Vitse, Branche, & Chamaillard, 2007; Prideaux, De Cruz, Ng, & Kamm, 2012) concluding that despite a broad panel of IBD-associated immune responses was published in the last decade (Table 1-1), their clinical usability is still very limited.

Alone serologic assessment of pANCA and ASCA has found clinical application to aid distinguishing CD and UC patients with ASCA+/pANCA- being indicative for CD and ASCA-/pANCA+ for UC. According to a meta-analysis of 60 studies with 7860 IBD subjects, CD patients can be distinguished from UC patients with 55 % sensitivity and 93 % specificity (Reese et al., 2006). In a recently published study referring to a commercial diagnostic test (Prometheus® IBD sgi™) the combination of serologic, genetic and inflammatory markers enhanced the diagnostic performance over single markers or the serologic marker panel alone (Plevy et al., 2013). Moreover, increased antibody titers were associated with a more complicated disease course in CD (Dubinsky et al., 2008; Ferrante et al., 2007; Mow et al., 2004; Tamboli, Doman, & Patel, 2011) and lower response to anti-TNF treatment in UC (Kevans et al., 2015; Nguyen, Nguyen, & Bechtold, 2015).

Table 1-1: Summary of published IBD antibody biomarkers

Marker	Description	Discovery approach	Reference
Autoantibodies			
ANCA **pANCA**	Anti-neutrophil cytoplasmic antibody	Immunofluorescence, fixed cell ELISA	(Rump et al., 1990; Saxon, Shanahan, Landers, Ganz, & Targan, 1990)
PAB **Anti-GP2**	Anti-pancreas antibodies (glycoprotein 2)	Immunofluorescence using a panel of tissues and IBD sera, antigen identification by 2D-Immunoblot and MALDI-TOF MS	(Roggenbuck et al., 2009; Stöcker et al., 1987)
Anti-GM-CSF	Antibodies against granulocyte-macrophage colony-stimulating factor (GM-CSF)	ELISA	(Han et al., 2009)
Antibodies against glycans			
ASCA	Anti- Saccharomyces cerevisiae antibodies (mannan)	ELISA against S. cerevisiae preparations	(Main et al., 1988)
Anti-L **Anti-C** **ALCA** **ACCA** **AMCA**	Antibodies against laminarin, chitin, laminaribioside, chitobioside and mannobioside	Glycan array using CD serum samples	(I. Dotan et al., 2006; Dotan, Altstock, Schwarz, & Dukler, 2006; Seow et al., 2009)
Antibodies against bacterial proteins			
Anti-OmpC	Antibodies against *E. coli* outer membrane protein C	Cross-reactivity of monoclonal pANCA with colonic bacteria using immunoblot	(Cohavy et al., 2000)
Anti-Cbir1, Anti-A4-Fla2 and Anti-FlaX	Antibodies against bacterial flagellin	SEREX using caecal metagenomic DNA of C3H/HeJBir mice	(Duck et al., 2007; Lodes et al., 2004)
Anti-I2	Antibodies against protein I2 protein sequence (tetR homologue) of *Pseudomonas fluorescens*,	Subtractive cloning in lamina propria mononuclear cells from the colonic mucosa of IBD patients	(Sutton, Kim, et al., 2000)
TCP-353	Antibodies against 8 aa peptide TCP-353	T7 phage display using cDNA of Caco-2 cells and CD patient serum	(Mitsuyama et al., 2011)
Anti-OmpW	Antibodies against a TonB-linked outer membrane protein of *Bacteroides caccae*	Cross-reactivity of monoclonal pANCA with colonic bacteria using immunoblot, SEREX using a *B. caccae* genomic expression library	(Cohavy et al., 2000; Wei et al., 2001)

Consequently, serologic markers have the potential to improve IBD diagnostics and disease management but cannot be used as a stand-alone diagnostic tool. Due to the low prevalence of known IBD-associated antibody responses, there is an urgent need for additional serologic markers in particular to classify subtypes of IBD and to predict disease course and drug response.

1.4.2 IBD mouse models

Numerous mouse models have been established to study molecular mechanisms underlying IBD and can mainly be grouped in three categories with defects in (1) epithelial barrier integrity, (2) innate immune mechanisms and (3) adaptive immune mechanisms. The use of IBD mouse models has recently been reviewed (Goyal, Rana, Ahlawat, Bijjem, & Kumar, 2014; Hörmannsperger, Schaubeck, & Haller, 2015; Wirtz & Neurath, 2007).

The model used in this work, transgenic $TNF^{\Delta ARE/+}$ mice, comprises a 69 bp deletion of the AU-rich elements (ARE) in the 3' UTR region of the TNFα gene conferring increased TNFα mRNA stability followed by elevated TNF levels (Kontoyiannis, Pasparakis, Pizarro, Cominelli, & Kollias, 1999). $TNF^{\Delta ARE/+}$ mice spontaneously develop inflammation of the terminal ileum. Ileal inflammation in this model was recently shown to be triggered by the gut microbiota, as germ-free mice did not develop inflammation (Schaubeck et al., 2015) and mice deficient in MyD88 (an adaptor protein essential for innate recognition of gut microbes) demonstrated attenuated signs of inflammation (Roulis et al., 2015). Moreover, shifts in the microbiota composition associated with inflammation were able to induce ileal inflammation when transferred to germ-free animals (Schaubeck et al., 2015). The involvement of both TNF and the gut microbiota are shared features between $TNF^{\Delta ARE/+}$ mice and human CD suggesting mechanistic similarities in the pathophysiology and making this mouse model well-suited to study host-microbiome interactions.

1.5 Objectives

The objective of this work was to develop a functional metaproteome display based on the ORFeome phage display technology and to identify novel gut microbiome-derived biomarker candidates for inflammatory bowel disease. After first providing the proof-of-concept in the transgenic $TNF^{\Delta ARE/+}$ mouse model of Crohn's like ileitis, novel biomarker candidates should be identified using clinical relevant IBD samples.

2 Materials and Methods

2.1 Material

2.1.1 Equipment

The equipment used in this study is listed in the following table:

Table 2-1: List of equipment used in the study

Equipment	Model	Manufacturer
96-well pipette	VIAFLO96	INTEGRA Biosciences, Zizers, Switzerland
Balances	Sartorius Analytic	Sartorius, Göttingen, Germany
Balances	Sartorius excellence	Sartorius, Göttingen, Germany
Blotting Machine	Trans-Blot SD	Bio-Rad Laboratories, München, Germany
Blotting Machine	Trans-Blot Turbo	Bio-Rad Laboratories, München, Germany
Bluetooth music streaming device	Soundweaver Bluetooth 4.0	EasyAcc, Hong Kong, HKG
Capillary electrophoresis	QIAxcel Advance	Qiagen, Hilden, Germany
Centrifuges	5415 D	Eppendorf, Hamburg, Germany
Centrifuges	Sorvall RC6 Plus	Thermo Scientific, Dreieich, Germany
Electrophoresis chamber	Biotechnology Model 40-0708	Peqlab, Erlangen, Germany
Electrophoresis chamber	Mini Protean III for SDS PAGE	Bio-Rad Laboratories, München, Germany
Electroporation device	MicroPulser™	Bio-Rad Laboratories, München, Germany
ELISA Reader	Sunrise	Tecan, Crailsheim, Germany
ELISA Reader	Epoch	BioTek Germany, Bad Friedrichshall, Germany
ELISA Washer	Columbus Pro	Tecan, Crailsheim, Germany
ELISA Washer	Columbus Plus	Tecan, Crailsheim, Germany
ELISA Washer	EL405	BioTek Germany, Bad Friedrichshall, Germany
Gel Documentation	ChemiDoc MP	Bio-Rad Laboratories, München, Germany
Incubators	BE 400	Memmert, Schwabach, Germany
Incubators	Vortemp 56	Labnet International, Edison, NJ USA
Incubators	Infors HT	Infors, Leipzig, Germany
Lab automation	QIAcube	Qiagen, Hilden, Germany
Laminar flow benches	Hera safe	Heraeus, Hanau, Germany
Laminar flow benches	MSCAdvantage	Thermo Scientific, Dreieich, Germany
Liquid handlig system	EL406	BioTek Germany, Bad Friedrichshall, Germany

Equipment	Model	Manufacturer
Microplate stacker	Biostack 3	BioTek Germany, Bad Friedrichshall, Germany
Microplate stacker	Biostack 4	BioTek Germany, Bad Friedrichshall, Germany
Photometer	NanoDrop ND1000	PeqLab, Erlangen, Germany
Photometer	Libra S11	Biochrom, Holliston, USA
Pipettes	Research	Eppendorf, Hamburg, Germany
Pipettes	Pipetman	Gilson, Middleton, USA
Pipettes	Biohit eLine e1200	Sartorius, Göttingen, Germany
Power Supplies	Electrophoresis Power Supply 301	Amersham Bioscience, Freiburg, Germany
Power Supplies	Electrophoresis Power Supply 601	Amersham Bioscience, Freiburg, Germany
Power Supplies	Electrophoresis Power Supply EPS-301	Amersham Bioscience, Freiburg, Germany
Power Supplies	PowerPac HC	Bio-Rad Laboratories, München, Germany
Rocker	GFL 3013	Omnilab Laborzentrum, Bremen, Germany
Rotors	SLA-300	Thermo Scientific, Dreieich, Germany
Rotors	SS34	Thermo Scientific, Dreieich, Germany
Rotors	F9S-4x1000y	Thermo Scientific, Dreieich, Germany
Rotors	A-4-81	Eppendorf, Hamburg, Germany
Thermocycler	DNAEngine	Bio-Rad Laboratories, München, Germany
Thermocycler	S1000 ThermalCycler	Bio-Rad Laboratories, München, Germany
Thermomixer	Compact	Eppendorf, Hamburg, Germany
Thermomixer	Comfort	Eppendorf, Hamburg, Germany
Water Installation	Arium 611	Sartorius, Göttingen, Germany

2.1.2 Consumables

The consumables used in this study are listed in the following table:

Table 2-2: List of consumables used in this study

Material	Manufacturer
Air-o-Seal hydrophobic Gas permeable seal	4titude, Dorking, UK
Amicon Ultra-0.5 mLCentrifugal Filters 30K	Merck Millipore, Darmstadt, Germany
Chelating Sepharose™ Fast Flow	GE Healthcare UK Ltd, Buckinghamshire, UK
Costar disposable serological pipettes (2 mL/5 mL/10 mL/25 mL)	Corning Inc., NY, USA
Costar MTP 96 Well Polystyrole plates	Corning Inc., Corning, NY, USA
Disposable Cuvettes	Brand, Wertheim, Germany
Electroporation cuvettes, GenePulser 0.1 cm	Bio-Rad Laboratories, München, Germany
Inoculation loops	VWR, Darmstadt, Germany
Microplate seal	4titude, Dorking, UK
MTP 384 Well Polystyrole	Greiner Bio-One, Frickenhausen, Germany
MTP 96 Well Polypropylene, U-shaped	Greiner-Bio-One, Frickenhausen, Germany
MTP cover foil AeraSeal	Excel Scientific, Victorville, CA, USA
MTP cover foil Sealplate	Excel Scientific, Victorville, CA, USA
Multiply-µStrip Pro 8-strip	Sarstedt, Nümbrecht, Germany
Petri dishes (7 cm, 15 cm)	Greiner Bio-One, Frickenhausen, Germany
Pipette tips	Sarstedt, Nümbrecht, Germany
Pipette tips, filtered	Greiner Bio-One, Frickenhausen, Germany
Pipette tips, filtered	Nerbe plus, Winsen, Germany
Polypropylene Centrifugation Tubes	Corning Inc., NY, USA
Polypropylene Columns	Qiagen, Hilden, Germany
PVDF membrane	Carl Roth, Karlsruhe, Germany
Reaction tube (1.5 mL, 2 mL)	Sarstedt, Nümbrecht, Germany
Reaction tube (15 mL, 50 mL)	Greiner Bio-One, Frickenhausen, Germany
Screw-top micro caps	Sarstedt, Nümbrecht, Germany
Spatulas	VWR, Darmstadt, Germany
Sterile filter (0.2 µm, 0.45 µm)	Sartorius, Göttingen, Germany

2.1.3 Chemicals

If not noted differently, all chemicals were obtained from the following companies: Carl Roth GmbH (Karlsruhe, Germany), Sigma-Aldrich Chemie GmbH (Steinheim, Germany), Merck KGaA (Darmstadt, Germany), ApplChem GmbH (Darmstadt, Germany), Roche

Diagnostics GmbH (Penzberg, Germany) and SERVA Electrophoresis GmbH (Heidelberg, Germany).

For all solutions, media and buffers prepared, cleaned and completely desalted water was used (Arium 611, Sartorius AG).

2.1.4 Enzymes and buffers

The enzymes and the corresponding buffers used in this study are listed in the following table:

Table 2-3: Enzymes and corresponding buffers used in this study

Enzyme	Manufacturer
DNA Polymerases	
GoTaq® G2 DNA polymerase	Promega, Madison, USA
5x Green GoTaq® Flexi Buffer	Promega, Madison, USA
$MgCl_2$ (25 mM)	Promega, Madison, USA
Phusion® High-Fidelity DNA Polymerase	Thermo Scientific, Dreieich, Germany
5x Phusion® HF Buffer	Thermo Scientific, Dreieich, Germany
Endonucleases	
NcoI-HF	New England BioLabs, Frankfurt a. M., Germany
NdeI	New England BioLabs, Frankfurt a. M., Germany
NheI-HF	New England BioLabs, Frankfurt a. M., Germany
NotI-HF	New England BioLabs, Frankfurt a. M., Germany
PmeI	New England BioLabs, Frankfurt a. M., Germany
CutSmart® Buffer	New England BioLabs, Frankfurt a. M., Germany
NEBuffer 3	New England BioLabs, Frankfurt a. M., Germany
Phosphatases	
Alkaline Phosphatase, Calf Intestinal (CIP)	New England BioLabs, Frankfurt a. M., Germany
DNA Ligases	
T4 DNA ligase	Promega, Madison, USA
Proteases	
Trypsin	Sigma-Aldrich Chemie GmbH, Steinheim, Germany

2.1.5 Antibodies

Antibodies were used as detecting reagents for ELISA and immunoblots as well as capture antibodies in ELISA. Secondary detection reagents were conjugated with horseradish peroxidase (HRP). The antibodies used in this study and the working dilutions are listed in the following table:

Table 2-4: Antibodies and working dilutions used in this study

Antibody	Conjugate	Dilution		Manufacturer
		ELISA	Immunostain	
Mouse anti-M13 (pVIII specific)	-	1:1,000 (capture)	-	Progen, 61097
Rabbit anti-M13 (pVIII specific)	-	1:5,000 (capture)	-	Pierce Biotechnology, PA1-46334
Goat anti-human (IgA, IgG, IgM)	-	1:500 (capture)	-	Dianova, 109-005-064
Goat anti-human (IgA, IgG, IgM)	HRP	1: 5,000 (detection)	-	Dianova, 109-036-064
Goat anti-mouse (IgA, IgG, IgM)	-	1:1,000 (capture)	-	Antikoerper-online, ABIN376851
Goat anti-mouse (IgA, IgG, IgM)	HRP	1:8,000 (detection)	1:4,000	Antikoerper-online, ABIN376237
Mouse anti-M13 (pVIII specific)	HRP	1:40,000 (detection)		GE, 27-9421-01

2.1.6 Commercial kits and markers

Commercial kits were used for the extraction and purification of nucleic acids (Table 2-5).

Table 2-5: Kits used for nucleic acid purification

Kit	Manufacturer
NucleoSpin Gel and PCR Clean-Up	Macherey-Nagel, Düren, Germany
peqGOLD MINIprep Kit I	Peqlab, Erlangen, Germany

DNA and protein markers were used as reference agarose gel- and capillary electrophoresis as well as in SDS-PAGE and immunoblots (Table 2-6).

Table 2-6:DNA and protein size standards used in this study

Marker	Manufacturer
DNA	
GeneRuler 1 kb Plus Ladder	Thermo Scientific, Dreieich, Germany
QX DNA Size Marker 100 bp - 2.5 kb	Qiagen, Hilden, Germany
QX Alignment Marker 15 bp/3 kb	Qiagen, Hilden, Germany
Protein	
Precision Plus Protein All Blue	Bio-Rad, München, Germany

Enzymatic kits were used for DNA amplification (MDA) and blunting/phosphorylation of DNA fragments (Table 2-7).

Table 2-7: Enzymatic kits used in this study

Kit	Manufacturer
Illustra RTG GenomiPhi V3	GE Healthcare UK Ltd, Buckinghamshire, UK
Fast DNA End Repair Kit	Thermo Scientific, Dreieich, Germany

2.1.7 Bacterial strains and bacteriophages

Bacterial strains and bacteriophages used in this study are listed below:

Table 2-8: List of bacterial strains and bacteriophages used in this study

Strain	Genotype	Application	Source
E. coli XL1-Blue-MRF'	Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1* gyrA96 relA1 lac *[F'* proAB lacI^{q}ZΔM15 *Tn*10 *(Tet^r)]*	Cloning	Stratagene, La Jolla, CA, USA
E. coli TOP10F'	F'[l*acI^q* Tn*10*(tet^R)] *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 deoR *nup*G recA1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1 λ$^-$	Library construction, panning	Invitrogen, Carlsbad, CA, USA
E. coli SS320 (MC1061 F')	F'[*proAB+lacIqlacZ*ΔM15 *Tn10* (tet^r)]*hsdR mcrB araD139 Δ(araABC-leu)*7679 Δ*lac*X74 *galUgalK rpsL thi*	Library construction	Lucigen, Middleton, WI USA
E. coli BLR(DE3)	F$^-$ *ompT hsdS*$_B$(r$_B^-$ m$_B^-$) *gal dcm* (DE3) Δ(srl-recA)306::Tn*10* (Tet^R)	Protein expression	Merck Biosciences Ltd, Nottingham, UK (former Novagen)
Hyperphage M13K07ΔgIII	-	Phage production	(Rondot et al., 2001)
M13K07	-	Serum preclearance	(Vieira & Messing, 1987)

2.1.8 Murine specimens

All murine samples used in this work were provided by Dr. Thomas Clavel (TU München, Germany) and derived from following mouse strains:

Table 2-9: List of mouse strains used in this study

Mouse strain	Description	Reference
C57BL/6N	Wildtype	
$TNF^{\Delta ARE/+}$	Mouse model of experimental ileitis, 69 bp deletion of AU-rich elements in 3' UTR of TNFα gene	(Kontoyiannis et al., 1999)

Animal use was approved by "Landratsamt Freising" (animal welfare authorization no. 32 568), following the guidelines of the "Deutsches Tierschutzgesetz" (German Animal Welfare Act) and the "Deutsche Tierschutz-Versuchstierverordnung" (German Animal Welfare of Experimental Animals Regulation) under supervision of a veterinarian and an animal welfare officer. Mice were sacrificed for scientific purposes only and were not included in any specific treatment protocols.

All mice were housed in the mouse facility at the Life Sciences faculty of the Technische Universität München under conventional or specific pathogen free (SPF) conditions with a 12 h light/dark cycle at 24-26 °C. All mice were fed a standard diet (R/M-H, Ssniff, Soest, Germany) *ad libitum* and were sacrificed by CO_2 inhalation at the indicated age (4 to 18 weeks).

Histologic scoring was performed using formalin fixed and paraffin embedded tissue sections of the distal ileum and proximal colon from $TNF^{\Delta ARE/+}$ and WT mice stained with hematoxylin and eosin (H&E) and scored in a blinded manner by assessing lamina propria mononuclear cell infiltration, crypt hyperplasia, goblet cell depletion and architectural distortion. Gut inflammation was assessed by scoring from 0 (non-inflamed) to 12 (highly inflamed) per section as described previously (Katakura et al., 2005).

2.1.8.1 Metagenomic DNA

Metagenomic DNA was isolated from the caecal content of $TNF^{\Delta ARE/+}$ mice (Technische Universität München). Samples were stored at -20 °C.

Table 2-10: Gut microbiome-derived metagenomic DNA from mice for library construction

Mouse ID	Library	Genotype	Housing	Age (weeks)
4643	1	$TNF^{\Delta ARE/+}$	Conv.	8
4647	2	$TNF^{\Delta ARE/+}$	Conv.	8
4651	3	$TNF^{\Delta ARE/+}$	Conv.	8

2.1.8.2 Blood samples

Blood was collected from the vena cava of sacrificed mice using tubes containing EDTA. The cells were separated by centrifugation and plasma (Table 2-11) stored at -80 °C for long-term or short-term storage at 4 °C.

Table 2-11: Mouse blood samples used in this study

Mouse/Cohort	n of samples	Genotype	Housing	Age
4643	1	$TNF^{\Delta ARE/+}$	Conv.	8
4647	1	$TNF^{\Delta ARE/+}$	Conv.	8
4651	1	$TNF^{\Delta ARE/+}$	Conv.	8
Candidate validation	17	$TNF^{\Delta ARE/+}$	Conv.	18
Candidate validation	19	wildtype	Conv.	18
Reactivity SPF	20	$TNF^{\Delta ARE/+}$	SPF	18
Reactivity SPF	20	wildtype	SPF	18
Time series	50 (10 of each age)	$TNF^{\Delta ARE/+}$	Conv.	4, 6, 8, 12, 18
Time series	49 (10 of each age, 9 at 18 weeks)	wildtype	Conv.	4, 6, 8, 12, 18

2.1.9 Human specimens

2.1.9.1 Metagenomic DNA

Human gut metagenomic DNA was provided from different sources (Table 2-12).

Table 2-12: Gut microbiome-derived metagenomic DNA from IBD patients for library construction

Library	Source	Provided by
1	CD, Colon, lumen, non-inflamed	Patricia Lepage, INRA, Joey-en-Josas, France
2	CD, Colon, lumen, non-inflamed	
3	CD, Caecum, lumen, non-inflamed	
4	CD, remission, fecal sample	Dr. Harry Sokol, St. Antoine Hospital, Paris, France
5	CD, remission, fecal sample	
6	UC, remission, fecal sample	
7	UC, remission, fecal sample	

2.1.9.2 Blood samples

Human sera used in this study was provided from different sources (Table 2-13). Storage was at -80 °C for long-term of 4 °C for short-term.

Table 2-13: Human blood samples used in this study

Cohort	n of samples	Phenotype	Description	Provided by
IBD cohort	50	25 CD/25 UC	Serum of IBD patients. Unknown ASCA/ANCA status.	Dr. Harry Sokol, St. Antoine Hospital, Paris, France
Preliminary control cohort	9	healthy	Healthy males. Exclusion criteria: among others chronic diseases (e.g. inflammatory bowel diseases) (Lagkouvardos et al., 2015)	Dr. Thomas Clavel, TU München, Germany
Control cohort	55	healthy	Age matched control sera. Exclusion criteria: any history of gastrointestinal disorder Sampled for the VID (*Virtuelles Institut Diabetes*) study	Prof. Dr. Hans Hauner and Dr. Thomas Skurk, TU München, Germany

2.1.10 Plasmids and oligonucleotides

In this study, different plasmids were used as backbones for phage display library construction or recombinant protein expression and are listed in the following table:

Table 2-14: Plasmid used in this study

Plasmid	Description	Reference
pHORF3	Phagemid, coding for oligopeptide:pIII fusion protein, library vector for metaproteome phage display libraries	(Kügler et al., 2008)
pET21a(+)	*E. coli* expression vector for cytoplasmic protein expression (T7 promoter)	Merck Biosciences Ltd, Nottingham, UK (former Novagen)
pET21a(+)-pelB	*E. coli* expression vector for periplasmic protein expression (T7 promoter)	TU Braunschweig, Dep. of Biotechnology, Braunschweig, Germany

Oligonucleotides were used for colony PCR or PCR primers to amplify coding sequences. Oligonucleotide use this work are listed in the following table:

Table 2-15: Oligonucleotide used in this study

ID	DNA sequence (5'→3')	Fwd./ rev.	Description
125	GGCTCGTATGTTGTGTGG	Fwd.	CPCR pHORF3 Sequencing pHORF3
11	CTAAAGTTTTGTCGTCTTTCC	Rev.	CPCR pHORF3
152	GAGCGGATAACAATTCCCC	Fwd.	CPCR pET21a(+), pET21a(+)-pelB Sequencing pET21a(+), pET21a(+)-pelB
153	GCAGCCAACTCAGCTTCC	Rev	CPCR pET21a(+), pET21a(+)-pelB Sequencing pET21a(+), pET21a(+)-pelB
2421	X-GCTCAGCCGGCGATGG	Fwd.	NGS pHORF3
2422	X-CAGCTCTGATATCTTTGGATCCC	Rev.	NGS pHORF3
2976	GATCCCATGGCTTGGAAGCCTGATCCGTCTTAC	Fwd.	JOZ240-H2 domain, NcoI
2977	GATCGCGGCCGCCTTGTATCTTCCCGCTGTTCTAC	Rev.	JOZ240-H2 domain, NotI
2978	GATCCCATGGCTGATTCCCAGTCTATGAACGG	Fwd.	JOZ241-A4 domain, NcoI
2979	GATCGCGGCCGCGGGAATATCCGCGATTG	Rev.	JOZ241-A4 domain, NotI
2980	GATCCCATGGCTATTTTGGCAACAGATGAAGATC	Fwd.	JOZ241-C4 domain, NcoI
2981	GATCGCGGCCGCCGCTTTTTTGCCCACTTC	Rev.	JOZ241-C4 domain, NotI
2982	GATCCCATGGCTTGGGATGCCGTGGCC	Fwd.	JOZ242-C9 domain, NcoI
2983	GATCGCGGCCGCAAAATAGAACGGCCTGCGAC	Rev.	JOZ242-C9 domain, NotI
2984	GATCCCATGGCTTCCAGGGACTCCTTGAGGAG	Fwd.	JOZ242-G4 domain, NcoI
2985	ATCGCGGCCGCTGCCGCCAAGAAGGC	Rev.	JOZ242-G4 domain, NotI
2986	GACTCCATGGCTTTGTCCAACATATCTCGGGATG	Fwd.	JOZ243-E9 domain, NcoI
2987	GATCGCGGCCGCACTGGGAATATCCGCGATTG	Rev.	JOZ243-E9 domain, NotI
2988	GATCCCATGGCTATGGCAACTGAGAACGATGATTC	Fwd.	JOZ241-A4 full-length, NcoI
2989	GATCGCGGCCGCTTCGCCAGCAGTACCTGTATCTG	Rev.	JOZ241-A4 full-length, NotI

ID	DNA sequence (5'→3')	Fwd./ rev.	Description
2990	GATCCATATGGCAACTGAGAACGATGATTC	Fwd.	JOZ241-A4 full-length, NdeI
2991	GATCCCATGGCTATGAAAACACTCATTAAAAATGTGC	Fwd.	JOZ241-C4 full-length, NcoI
2992	GATCGCGGCCGCGAGAGCAGCTCTTTTTTTAGCTTC	Rev.	JOZ241-C4 full-length, NotI
2993	GATCCATATGAAAACACTCATTAAAAATGTGC	Fwd.	JOZ241-C4 full-length, NdeI
2994	GATCGCTAGCATGTGCAGGAGATTCGCG	Fwd.	JOZ242-C9 full-length, NheI
2995	GATCGCGGCCGCAAACAGACGCATGGGTTCG	Rev.	JOZ42-C9 full-length, NotI
2996	GATCCATATGTGCAGGAGATTCGCGC	Fwd.	JOZ242-C9 full-length, NdeI
3001	GATCCCATGGCTATGGTTAAATTACCGCCGCTG	Fwd.	JOZ191-G8 full-length, NcoI
3002	GATCGCGGCCGCCTCAGCCAGAAAAAGCTCCAG	Rev	JOZ191-G8 full-length, NotI
3003	GATCCATATGGTTAAATTACCGCCGCTG	Fwd.	JOZ191-G8 full-length, NdeI
3026	GATCCCATGGCTATGCAGGATATTGCGATTGAGG	Fwd.	JOZ243-F9, NcoI
3027	GATCGCGGCCGCTACCGTTTCCGGCAGC	Rev.	JOZ243-F9 domain, NotI
3028	GATCGCGGCCGCCTTCACCCACTTTTGAGCGTTC	Rev.	JOZ243-F9 full-length, NotI
3085	GATCCCATGGCTTGGAAGCCTGATCCGTCTTAC	Fwd.	JOZ240-H2 domain, NdeI
3086	GATCCATATGGATTCCCAGTCTATGAACGG	Fwd.	JOZ241-A4 domain, NdeI
3087	GATCCATATGATTTTGGCAACAGATGAAGATC	Fwd.	JOZ241-C4 domain, NdeI
3088	GATCCATATGTGGGATGCCGTGGCC	Fwd.	JOZ242-C9 domain, NdeI
3096	GATCCATATGTCCAGGGACTCCTTGAGGAGCG	Fwd.	JOZ242-G4 domain, NdeI
3090	GATCCATATGTTGTCCAACATATCTCGGGATG	Fwd.	JOZ243-E9 domain, NdeI
3100	GATCCATATGCAGGATATTGCGATTGAGG	Fwd.	JOZ243-F9 domain, NdeI

2.1.11 Peptides

All peptides used in this study (Table 2-16) were purchased from Peps4LS (Heidelberg, Germany) either as raw product or HPLC purified. N-terminal biotinylated peptides were purchased with a short PEG-spacer (eBio) and glycine-serine linker.

Table 2-16: Peptides used in this study

Peptide	Sequence	Purification
JOZ156-H5	eBio-GGSGGS-WDFMGNFWNFNSYRDFDFPR	Raw product
JOZ158-E11	eBio-GGSGGS-VFWPADGYEPGHGQPSFDKQFARDW LKENDGHDWTLPQEIVE	Raw product
JOZ158-G8	eBio-GGSGGS-VSNDLNALLETNISAGTRTTQVSMSYF GEMLMSHIADNEFLSGSEDEKAAHANALA	Raw product
JOZ191-E5	eBio-GGSGGS-FSGPSHRTAPVPSGGWGTQVTQGR	Raw product
JOZ191-G8	eBio-GGSGGS-IEPNTLFGSRPPVLPDDDALWDIFEQG HQLLTA	Raw product
JOZ191-G8	IEPNTLFGSRPPVLPDDDALWDIFEQGHQLLTC	HPLC
JOZ192-A10	eBio-GGSGGS-YIVSAPSGAGKSSLIQALLKTQ	Raw product
JOZ193-A9	eBio-GGSGGS-VVEGTSGSTKPVLGANVK	Raw product
JOZ240-H2	WKPDPSYLCTCSEYAASSAPMPASCASTSSACRTAGR YKC	Raw product
JOZ243-F12	SALTAAASAERC	HPLC
JOZ243-F9	MQDIAIEDTPSSFAATVDDVQNSSNSMPDNPNATLPET VC	HPLC

2.1.12 Media and supplements

If not noted otherwise*, E. coli* was generally cultured in fully buffered two times yeast-tryptone media (2xYT) (Table 2-17). For solid media 12 g agar per liter was added.

Table 2-17: Recipe for 2x YT medium

2x YT medium	
Bacto-Yeast Extract	1.0 % (w/v)
Bacto-Tryptone	1.6 % (w/v)
NaCl	0.5 % (w/v)

SOC medium (Table 2-18) was used for regeneration of *E. coli* cells after transformation.

Table 2-18: Recipe for SOC medium

SOC medium	
Bacto-Yeast Extract	0.5 % (w/v)
Bacto-Tryptone	2.0 % (w/v)
NaCl	0.05 % (w/v)
Glucose	1.8 % (w/v)

All media were sterilized by autoclaving (121°C, 20 min and excess pressure 1 bar). For solid media Agar-Agar was added before autoclaving.

Media supplements (Table 2-19) were prepared as concentrated stock solutions and autoclaved or filter sterilized (0.2 µm pore size) if heat sensitive. Addition of supplements

to the media was performed under sterile conditions. For solid media supplements were added after media had reached a temperature below 55°C.

Table 2-19: Medium supplements used in this study

Supplement	Stock solution		Final concentration	
Ampicillin	100	mg/mL	100	µg/mL
Kanamycin	50	mg/mL	50	µg/mL
Tetracycline	10	mg/mL	20	µg/mL
Glucose	2	M	100	mM
IPTG	1	M	1	mM

Supplemented 2xYT media used in this study are listed in the following table:

Table 2-20: List of media used in this study

Medium	Supplements
2x YT-A	100 µg/mL ampicillin
2x YT-GA	100 mM glucose, 100 µg/mL ampicillin
2x YT-AK	100 µg/mL ampicillin, 50 µg/mL kanamycin
2x YT-T	20 µg/mL tetracycline

2.1.13 Buffers and solutions

Buffers used in this studies are given in the following table:

Table 2-21: List and composition of buffers and solutions used in this study

3,3-diaminobenzidine (DAB) stock solution				
3,3'-diaminobenzidine	2.5	%	(w/v)	in PBS
Acrylamide Mix				
Acrylamide	30	%	(w/v)	
Bisacrylamide	0.8	%	(w/v)	
Agarose Gel				in TAE
Agarose	1.0	%	(w/v)	Buffer
Ammonium persulfate (APS-) Solution				
APS	10	%	(w/v)	
Blotting Buffer				
Tris	25	mM		
Glycine	192	mM		
BPBST				
Bovine serum albumin (BSA)	2	%	(w/v)	in PBST (0.1 %)
Coomassie blue destaining solution				
Acetic acid	10	%	(v/v)	

Coomassie blue staining solution				
Coomassie Brilliant Blue R250	0.05	%	(w/v)	
Acetic acid	10	%	(v/v)	
2-Propanol	2.5	%	(v/v)	
DAB substrate solution				
$CoCl_2$	0.02	%	(w/v)	In PBS
DAB	0.05	%	(w/v)	
Hydrogen peroxide	0.03	%		
ELISA washing buffer				
Tween20	0.05	%	(w/v)	in H_2O
Ethidium Bromide Solution				
Ethidium Bromide	0.1	%	(w/v)	
IMAC binding buffer				
Na_2HPO_4	20	mM		
NaH_2PO_4	20	mM		
NaCl	500	mM		
Imidazole	10	mM		
IMAC Elution buffer				
EDTA	100	mM		in PBS
MPBS				
Skim Milk Powder	2	%	(w/v)	in PBS
MPBST				
Skim Milk Powder	2	%	(w/v)	in PBST (0.1 %)
OS buffer				
$MgSO_4$	5	mM		
PBST (0.05 - 0.1 %)				
Tween20	0.05 - 0.1	%	(v/v)	in PBS
PE buffer, pH 8.0				
Sucrose	20	%	(w/v)	
Tris	50	mM		
EDTA	1	mM		
Phage dilution buffer				
Tris-HCl pH7.5	10	mM		
NaCl	20	mM		
EDTA	2	mM		
Phage elution buffer				
Trypsin	10	µg/mL		in PBS
Phage precipitation buffer				
PEG 6000	20	%	(w/v)	
NaCl	2.5	M		
Phosphate buffered saline (PBS)				
NaCl	0.8	%	(w/v)	
KCl	0.02	%	(w/v)	
Na_2HPO_4 x 2 H_2O	0.144	%	(w/v)	
KH_2HPO_4	0.024	%	(w/v)	
SDS sample (Laemmli) buffer (5x)				
SDS	10	%	(w/v)	
Glycerol	50	%	(w/v)	
Bromophenol Blue	0.02	%	(w/v)	
β-Mercaptoethanol	15	%	(v/v)	

SDS Solution			
SDS	10	%	(w/v)
SDS-PAGE Running Buffer			
Tris	25	mM	
Glycine	192	mM	
SDS	0.1	%	(w/v)
TAE Buffer			
Tris-HCl	4	mM	
Acetic acid	2	mM	
EDTA	1	mM	
TMB Reagent (20A:1B)			
Solution A, pH 4.1			
Potassium Citrate	30	mM	
Citric acid	50	mM	
Solution B			
3,3',5,5'-Tetramethylbenzidine	1	mM	
Acetone	10	%	(v/v)
Ethanol	90	%	(v/v)
Hydrogen peroxide	0.3	%	(v/v)

2.1.14 Software and databases

The software and databases used for this work are listed in the following table:

Table 2-22: Software and databases used in this study

Software/database	Application	Manufacturer/Source
Vector NTI Advance 11.5	Vector maps DNA and protein alignments DNA sequence processing	Invitrogen, Carlsbad, CA, US
Geneious 4.8.5	Vector maps DNA and protein alignments In silico cloning	Biomatters Ltd, Auckland, New Zealand
NCBI	Literature research, sequence comparisons (BLAST), gene and protein sequence information	http://www.ncbi.nlm.nih.gov/
UniProt	Protein sequence database	http://www.uniprot.org/
Graphpad Prism	Data presentation and statistical analysis	
QIAxcel ScreenGel	Capillary electrophoresis analysis	Qiagen, Hilden, Germany
ImageLab	Documentation of gelelectrophoresis	Bio-Rad
Illustrator	Figure illustration	Adobe
Endnote	Literature database Citing tool	Thomson Reuters, NYC, USA
Gen5 2.08	Readout of ELISA plates	BioTek Instruments, USA
Liquid Handling Control 2.17	Controlling of liquid handling system EL406	BioTek Instruments, USA

2.2 Methods

2.2.1 Bacterial cultivation techniques

2.2.1.1 Plating

After transformation or phage infection of *E. coli*, cell suspensions were plated on a 2xYT agar plates supplemented with 0.1 M glucose and the respective antibiotic with a spatula or glass beads. Incubation was performed at 37°C overnight.

2.2.1.2 Liquid cultures

Starter cultures were prepared in 5 mL or 30 mL scale in 2x YT media supplemented with the respective antibiotic and optional 0.1 M glucose. Inoculation was performed either from a single colony taken from a plate or directly from a glycerol stock. Cultures were incubated at 37 °C and 250 rpm for 16 to 18 h.

Main cultures of volumes between 30 and 400 mL were inoculated from the starter culture with OD_{600} adjusted to 0.05 to 0.1. Composition of the medium was generally set up as for the starter culture. Incubation of main cultures was performed at 37 °C and 250 rpm in a shaking incubator logarithmic growth (OD_{600} = 0.5) was reached.

2.2.1.3 Storage of microbial cultures

Bacterial cultures were stored short-term on solid media for up to two weeks at 4 °C. For long-term storage of bacterial cultures, liquid cultures were supplemented with 20 % (v/v) glycerol and stored at -80 °C.

2.2.2 Transformation of bacteria

2.2.2.1 Transformation by heat shock

For the transformation of *E. coli* cells after ligation, 20 µL of frozen chemically competent cells were carefully thawed on ice. The cells were mixed with 10 µL of the ligation and incubated on ice for 20 min. The cells were incubated at exactly 42 °C for 60 s and subsequently chilled for 2 min on ice. 1 mL pre-warmed SOC media was added and the cells were incubated for 1 h at 37 °C and 650 rpm. The cells were pelleted for 30 s at 16,100 xg and resuspended in 100 µL supernatant. The cell suspension was plated on 2x YT-G supplemented with the appropriate antibiotic and incubated at 37 °C overnight.

2.2.2.2 Transformation by electroporation

Electroporation was used for the transformation of electro-competent *E. coli* cells for metaproteome phage display library construction and recombinant protein expression.

When transforming cells using a previous ligation reaction, a buffer exchange of the DNA was performed 4 times with 500 µL H_2O using Amicon Ultra Centrifugal Filters (30K) (Merck Millipore).

Frozen electro-competent *E. coli* cells were thawed on ice and mixed with the DNA. The cell/DNA mixture was transferred to pre-chilled (-20 °C) electroporation cuvettes and transformation was performed at 1.7 kV (*E.coli* BLR(DE3)) or 1.8 kV (*E. coli* TOP10F' and SS320) (MicroPulser™, BioRad). Immediately after the electroporation pulse, 1 mL of pre-warmed (37 °C) SOC was added. After 60 min of incubation at 37 °C and 600 rpm the cells were plated on 2x YT agar containing the appropriate supplements.

2.2.2.3 Preparation of plasmid DNA from *E. coli*

Preparation of plasmid DNA was performed using the peqGOLD MINIprep Kit I (Peqlab, Erlangen) following the manufacturer's instructions. *E. coli* carrying the plasmid of interest was cultured in 5 mL 2xYT GA medium at 37°C and 250 rpm overnight. The cells were harvested by centrifugation at 16,100 xg for 10 min and plasmid DNA was isolated according to the supplier's instructions.

2.2.3 Phage handling

2.2.3.1 Production of phage particles

2.2.3.1.1 400 mL scale (oligopeptide phage libraries)

Glycerol stocks of the metaproteome phage display libraries (2.2.4.6) were thawed completely and used to inoculate 400 mL 2x YT-GA medium (OD_{600} < 0.1). The cells were cultivated at 37 °C and 250 rpm (Infors HT) until logarithmic growth (OD_{600} = 0.5) was reached. In order to complement the missing coat proteins and ensure enrichment of ORFs, 25 mL of the culture were infected with a 20-fold excess (2.5x 10^{11} colony forming units (cfu) of the M13K07ΔgIII helperphage "Hyperphage" for 30 min at 37 °C. The infected cells were incubated for another 30 min at 37 °C and 250 rpm (Infors HT) to express antibiotic resistance. The cells were pelleted (3,220 xg, 10 min) and subsequently resuspended in 400 mL 2x YT-AK.

Oligopeptide phage particles were produced for 24 h at 30 °C and 250 rpm (Infors HT). The cells were pelleted and phage particles were precipitated from the supernatant at 4 °C overnight after adding 1/5 volume phage precipitation buffer. The precipitated phage particles were pelleted for 1 h at 12,000 xg and 4 °C (Sorval Centrifuge RC5B Plus, Rotor F9S) and resuspended in 10 mL phage dilution buffer. In order to confer higher purity, the resuspended phage were filtered (Whatman syringe filter, 0.45 µm) and a second precipitation step was performed for 2 h at 4 °C with 1/5 volume phage precipitation buffer. The phage particles were pelleted for 30 min at 20,000 xg and 4 °C (Sorval Centrifuge RC5B Plus, Rotor SS34) and resuspended in 1 mL fresh phage dilution buffer. Remaining bacteria were pelleted for 2 min at 16,100 xg and supernatants containing the oligopeptide phage libraries were collected and stored at 4 °C.

2.2.3.1.2 30 mL scale (oligoclonal phage or monoclonal phage)

The production of 30 mL scale phage production was performed during phage amplification between panning rounds (2.2.5.5) or the production of monoclonal phage for characterization in immunoblot and ELISA (2.2.7.1 and 2.2.7.2.2).

Thirty mL of 2x YT-GA were inoculated (OD_{600} < 0.1) and cultivated at 37 °C and 250 rpm (Infors HT) until logarithmic growth phase (OD_{600} = 0.5) was reached. Helperphage infection was performed for 30 min at 37 °C using 5 mL of the culture and a 20-fold excess (5x 10^{10} cfu) of the M13K07ΔgIII helperphage “Hyperphage”. The infected cells were incubated at for another 30 min at 37 °C and 250 rpm (Infors HT) to express antibiotics resistance. The cells were pelleted (3,220 xg, 10 min) and subsequently resuspended in 30 mL 2x YT-AK following phage production at 30 °C and 250 rpm overnight.

Amplified phage were precipitated (1 h on ice) after adding 1/5 volume phage precipitation buffer. After pelleting (1 h at 4 °C, 3,220 xg), the oligopeptide phage were resuspended in 500 µL phage dilution buffer. Remaining bacteria were pelleted for 2 min at 16,100 xg, the amplified oligopeptide phage containing supernatant was collected and stored at 4 °C until used.

2.2.3.1.3 MTP scale (monoclonal phage)

For the screening of monoclonal phage after panning (2.2.7.2.1), phage particles were produced in 96-well microtiter plate (MTP) scale.

A 96-well MTP was supplemented with 180 µL 2x YT-GA medium, inoculated with single colonies bearing a phagemid and incubated at 34 °C and 800 rpm (Labnet Vortemp 56) overnight. For phage production, 175 µL 2x YT-GA per well were inoculated with 10 µL of the overnight culture and incubated for 2 h at 37 °C and 800 rpm to reach logarithmic growth. The cells were infected with 5 x 10^9 cfu "Hyperphage" (M13K07ΔgIII) for 30 min at 37 °C. To ensure antibiotic resistance, the cells were incubated for another 30 min at 37 °C and 800 rpm. In order to change the medium for phage production, the cells were pelleted for 10 min at 3,220 xg and resuspended in 180 µL 2x YT-AK followed by phage production at 30 °C and 800 rpm overnight. The cells were pelleted and the phage were precipitated (1 h on ice) after adding 40 µL phage precipitation buffer to 150 µL phage containing supernatant. The phage particles were pelleted for 1 h at 3,220 xg an 4 °C, the supernatant was discarded and the phage were resuspended in 150 µL phage dilution buffer. Remaining bacteria were pelleted for 10 min at 3,220 xg.

2.2.3.2 Determination of phage titers

Phage titers were determined for preparation of oligopeptide phage libraries and produced monoclonal oligopeptide phage as well as to monitor panning output.

Phage titers were determined by infection of *E. coli* cells. Ten-fold serial dilutions of the phage were prepared in PBS. For infection, 50 µL of an *E. coli* XL1 Blue MRF' or *E. coli* TOP10F' culture at logarithmic growth (OD_{600} = 0.5 in 2xYT-T) was briefly mixed with 10 µL of a phage dilution and incubated for 30 min at 37°C. After the infection, 3x 10 µL of each dilution were spotted on 2x YT-GA agar plates. The plates were incubated at 37 °C overnight and the colonies were counted the next day.

To exclude phage contamination of the used PBS, a control with 50 µL *E. coli* culture and PBS was treated equally to the infection with phage. Also, the used PBS and *E. coli* culture alone was spotted on 2x YT-GA agar plates to exclude bacterial contamination of the PBS or ampicillin resistance of the *E. coli* culture. Another 50 µL of the *E. coli* culture were infected with 10 µL of a phage solution with known titer as a positive control.

2.2.4 Molecular biological methods

2.2.4.1 Polymerase chain reaction (PCR)

2.2.4.1.1 Colony-PCR

Colony-PCR war used to determine insert rates and mean insert size of metaproteome phage display libraries and to confirm the correct insert size after cloning.

Table 2-23: Combinations of oligonucleotide primers for colony PCR

Plasmid	Oligonucleotide primer (fwd./rev.)
pHORF3	125/11
pET21a(+)	152/153
pET21a(+)-pelB	152/153

The composition of the PCR reaction (Table 2-24) and the temperature profile (Table 2-25) for colony PCRs are listed in the following tables.

Table 2-24: Standard composition of a colony PCR

Component	Volume (µl)	Final conc.
Template (Cell material from one colony)	-	
Primer *forward* [10 µM]	0.5	500 nM
Primer *reverse* [10 µM]	0.5	500 nM
dNTP mix [10 mM]	0.2	200 µM
5x Green GoTaq Reaction Buffer	2	1 x
$MgCl_2$ [25mM]	0.8	2 mM
GoTaq DNA polymerase (5 U/µL)	0.05	0.025 U/µL
dH_2O	5.95	
Total volume	10	

Table 2-25: Temperature profile of a colony PCR (GoTaq®)

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	120	1x
Denaturation	95	15	25x
Annealing	x	20	
Elongation	72	Dependent on insert size (1 min/kb)	
Final elongation	72	600	1x
Soak	16	∞	

2.2.4.1.2 Gene amplification

The genes coding for the protein domains and full-length proteins of selected biomarker candidates were amplified by PCR using as template plasmid DNA containing the coding

sequence and amplified metagenomic DNA, respectively (Table 2-26). For cloning into the destination vectors pET21a(+) and pET21a(+)-pelB, the appropriate restrictions sites were introduced into the genes by PCR.

Table 2-26: Combination of oligonucleotide primers for the amplification of coding sequences

Clone	Destination vector	Oligonucleotide primer (fwd./rev.)	Annealing temp. (°C)
JOZ191-G8 full-length	pET21a(+)	3003/3002	53
	pET21a(+)-pelB	3001/3002	53
JOZ240-H2 domain	pET21a(+)	3085/2977	55
	pET21a(+)-pelB	2976/2977	51
JOZ241-A4 domain	pET21a(+)	3086/2977	55
	pET21a(+)-pelB	2978/2979	51
JOZ241-A4 full-length	pET21a(+)	2990/2989	52
	pET21a(+)-pelB	2988/2989	52
JOZ241-C4 domain	pET21a(+)	3087/2981	55
	pET21a(+)-pelB	2980/2981	51
JOZ241-C4 full-length	pET21a(+)	2993/2992	52
	pET21a(+)-pelB	2991/2992	52
JOZ242-C9 domain	pET21a(+)	3088/2983	55
	pET21a(+)-pelB	2982/2983	51
JOZ242-C9 full-length	pET21a(+)	2996/2995	52
	pET21a(+)-pelB	2994/2995	52
JOZ242-G4 domain	pET21a(+)	3096/2985	55
	pET21a(+)-pelB	2984/2985	51
JOZ243-E9 domain	pET21a(+)	3090/2987	52
	pET21a(+)-pelB	2986/2987	52
JOZ243-F9 domain	pET21a(+)	3100/3027	55
	pET21a(+)-pelB	3026/3027	51
JOZ243-F9 full length	pET21a(+)	3100/3028	55
	pET21a(+)-pelB	3026/3028	51

The composition of PCR reactions (Table 2-27) and the temperature profile (Table 2-28) for gene amplification using Phusion® high-fidelity polymerase are listed in the following tables

Table 2-27: Standard composition of a PCR for gene amplification (Phusion®)

Component	Volume (µl)	Final conc.
Template Metagenomic DNA Plasmid DNA	 x x	 3000-5000 ng/mL 40 ng/mL
Primer *forward* [10 µM]	0.5	100 nM
Primer *reverse* [10 µM]	0.5	100 nM
dNTP mix [10 mM]	1	200 µM
5x Phusion® HF Buffer	10	1 x
Phusion® High-Fidelity DNA Polymerase (2 U/µL)	0.5	0.02 U/µL
dH_2O	37.5-x	
Total volume	50	

Table 2-28: Temperature profile of a standard PCR for gene amplification

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	120	1x
Denaturation	95	15	30x
Annealing	56	15	
Elongation	72	Dependent on insert size (15-20 s/kb)	
Final elongation	72	600	1x
Soak	16	∞	

The PCR product was purified directly or after gel extraction using the NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel).

2.2.4.2 Multiple displacement amplification (MDA)

Metagenomic DNA was amplified by multiple displacement amplification according to the manufacturer's instructions (illustra RTG GenomiPhi V3 DNA Amplification Kit, GE Healthcare). Briefly, 100 ng metagenomic template DNA was used per 20 µL reaction. The template DNA was denatured for 3 min at 95 °C in denaturation buffer and subsequently chilled to 4 °C. The template DNA was transferred to the lyophilized reaction cake and incubated for 2 h at 30°C. Each amplification was performed in triplicates and the amplified DNA was pooled. Amplified DNA concentration was measured using a

spectrophotometer with a control reaction as reference. Integrity of amplified DNA was analyzed on a 1 % (w/v) agarose gel.

2.2.4.3 Electrophoresis

2.2.4.3.1 Agarose gel electrophoresis

Agarose gels (1 % (w/v)) were supplemented with 10 µL ethidium bromide solution per 40 mL of gel and the separation was performed at a constant voltage of 130 V for 20 min in TAE buffer. Documentation of the DNA fragments was performed under UV radiation (λ = 312 nm) using a camera system (ChemiDoc MP, Bio-Rad).

2.2.4.3.2 Capillary electrophoresis

Colony PCR analysis to determine insert rates and mean insert sizes of constructed libraries was performed by capillary electrophoresis (QIAxcel Advance) using a QIAxcel DNA Fast Analysis cartridge. Capillary electrophoresis was analyzed using the QIAxcel ScreenGel software and exported as virtual gel image. Minimal volumes of 15 µL were used. If PCR was performed in smaller volumes, the volume was increased to 15 µL by diluting with QX DNA Dilution Buffer (Qiagen).

2.2.4.4 DNA sequencing

2.2.4.4.1 Sanger sequencing

All DNA sequences were obtained from GATC Biotech GmbH (Konstanz, Germany) or Seqlab Sequence Laboratories GmbH (Göttingen, Germany) after providing purified plasmid DNA or an *E. coli* liquid culture containing the plasmid DNA.

2.2.4.4.2 Next generation sequencing (NGS)

NGS analysis was performed at the sequencing facility of ZIEL, TU München and raw data processing was performed by Dr. Ilias Lagkouvardos (TU München, Germany).

Briefly, nucleotide libraries were constructed by PCR (25 cycles) adding adaptor and barcode sequences using primer 2421 and 2422 (Table 2-15). They were purified and sequenced in paired-modus using a MiSeq Illumina sequencer. Reads were merged, cleaned from primer and vector sequences, and those with more than 0.1 % expected error rate were discarded using USEARCH v8.1.18 (Edgar, 2010). When different samples or replicates were compared, sequences were first dereplicated to allow comparison of unique reads across samples.

2.2.4.5 DNA Cloning

2.2.4.5.1 DNA fragmentation for library cloning

DNA was diluted in 2 mL H_2O and fragmented by sonication (6x 2 min, 50 % intensity, MS72 sonotrode, HD2200 Sonopuls). Fragmented DNA was analyzed on 1 % agarose gels to ensure fragment sizes between 200 and 1500 bp. The fragmented DNA was concentrated using Amicon Ultra Centrifugal Filters (30K) (Merck Millipore).

2.2.4.5.2 Removal of cohesive ends and DNA phosphorylation

Cohesive ends were blunted and blunt ends were phosphorylated according to manufacturer's instructions (Fast DNA End Repair Kit, Thermo Scientific). Briefly, 5 µg fragmented DNA in a total reaction volume of 50 µL were incubated for 16 min at 20 °C. The DNA was then purified using a spin column according to the manufacturer's instructions (NucleoSpin Gel and PCR Clean-UP Kit, Macherey-Nagel).

2.2.4.5.3 Endonuclease digest

Library vector pHORF3 was digested in one batch and used for all library constructions. Plasmid DNA was linearized using 5 U PmeI/µg plasmid DNA in CutSmart buffer (New England Biolabs) and a total reaction volume of 200 µL for 1h at 37 °C.

PCR fragments and plasmid DNA for the construction of expression vectors were digested with the appropriate enzymes using the buffer recommended by the manufacturer. Restriction digest was performed in 20-50 µL reaction volumes using up to 3 µg DNA and 20 U endonuclease per reaction. If available, NEB high fidelity (HF) enzymes were used. All reactions were performed at 37 °C for 1-2 h.

2.2.4.5.4 Dephosphorylation of linearized vector DNA

Vector DNA was dephosphorylated in order to prevent religation events. Dephosphorylation was usually performed in parallel with endonuclease digest by adding 2x 5U CIP (New England Biolabs) to the restriction digest and incubation for 30 min at 37°C each.

In order to increase dephosphorylation efficiency of blunt end linearized pHORF3 library plasmid, purified linearized plasmid DNA was incubated with 2 U CIP/µg DNA at a DNA concentration of 60 ng/µL using NEB3 buffer (New England Biolabs) and a total reaction volume of 150 µL for 1h at 50 °C.

2.2.4.5.5 Purification of DNA fragments

DNA fragments and PCR products were purified either directly or after extraction from agarose gels using the NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) according to the manufacturer's instructions.

2.2.4.5.6 Spectrophotometric determination of DNA concentrations

DNA concentrations were determined using a spectrophotometer (NanoDrop ND1000, λ = 260 nm).

2.2.4.6 Construction of metaproteome phage display libraries

The libraries were constructed using 1,400 ng fragmented, blunted and phosphorylated metagenomic DNA and 1000 ng PmeI (New England Biolabs) linearized and dephosphorylated pHORF3 library vector. The ligation was performed in a total reaction volume of 100 µL using 10 U T4 DNA Ligase (Promega) for 16 h at 16 °C and ligase was heat inactivated for 10 min at 65 °C. After buffer exchange, the ligation was split into 4 separate transformations and 25 µL of electro-competent *E. coli* (TOP10F' or SS320) were transformed by electroporation as described above (2.2.2.2). Transformation rates were determined by plating dilutions on 2 x YT-GA agar. Remaining cells were plated on 2 x YT-GA agar plates (25 x 25 cm) and incubated at 37 °C overnight. The cells were scraped using 20 mL of 2 x YT medium, flash frozen in liquid nitrogen and stored at -80 °C in 20 % (v/v) glycerol. Insert rates and mean insert sizes were determined by colony PCR (2.2.4.1.1).and capillary electrophoresis (2.2.4.3.2) of randomly analyzed colonies (n = 20 per transformation)

ORF enrichment and phage packaging of the libraries was performed as described above (2.2.3.1.1)

2.2.4.7 Construction of expression vectors for recombinant protein expression

Construction of the expression vectors for recombinant protein expression was performed using ligation reactions of 50 ng of linearized and dephosphorylated vector (pET21a(+) or pET21a(+)-pelB) and 3-5-fold molecular excess of insert.

$$Insert\ [ng] = \frac{ng\ Vector \times kb\ Insert}{kb\ Vector} \times \mathrm{n}\ (-\text{fold excess of insert})$$

Ligations were performed in total volumes of 10 µL using 1 U of T4 DNA Ligase (Promega). The samples were incubated for 16 h at 16 °C and inactivated for 10 min at 65°C.

The ligation reaction was used to transform chemically competent *E. coli* XL1 Blue MRF' by heat shock transformation.

2.2.5 Panning of metaproteome phage display libraries

2.2.5.1 Preclearance of serum from phage binders

Panning allows the selection of displayed oligopeptides based on specific affinity binding to serum antibodies. In order to reduce background signals, the serum used for panning was precleared from antibodies reactive with phage particles alone. Therefore, two wells of a 96-well ELISA plate (Costar MTP 96 Well Polystyrole, Corning) were coated with 4x 10^{10} cfu Hyperphage in 300 µL PBS. The wells were blocked with 350 µL MPBST per well. The blocking solution was removed and the sera (1:500 dilution in 300 µL MPBST) were incubated for 2x 1 h on immobilized Hyperphage to preclear anti-phage serum antibodies.

2.2.5.2 Immobilization of serum antibodies

The precleared serum was transferred into 2 blocked wells (150 µL each) previously coated with goat anti-mouse (IgA, IgG, IgM) (antikoerper-online, ABIN376851, 1:1,000 in PBS)) and goat anti-human (IgA, IgG, IgM) (109-005-064, Dianova, 1:500 in PBS) antibody, respectively, and incubated for 2 h at room temperature). Excess serum antibodies and other serum proteins were removed by washing 3 times with PBST (0.05%).

2.2.5.3 Preclearance of oligopeptide phage library from undesired clones (optional)

An optional preclearance step of the libraries to remove universal immunoglobulin-binding proteins and proteins non-specific to IBD (matched panning approach, human) was partially included. The libraries were incubated on control serum antibodies previously captured using an immobilized goat anti-human (IgA, IgG, IgM) antibody (1:500 in PBS,109-005-064, Dianova) for 2 h at room temperature.

2.2.5.4 Selection of specifically bound oligopeptides

For the selection of specifically bound oligopeptides, two to three rounds of panning were performed using 2x 10^{10} cfu oligopeptide phage library (in 150 µL MPBST) as input phage in round one and 100 µL of amplified oligopeptide phage (accounting for 3x 10^{11} cfu) as input in round two and three.

The oligopeptide phage were incubated with the immobilized serum antibodies for 2 h at room temperature and unbound oligopeptide phage were removed by stringent washing (10, 20, 30 times with PBST (0.05 %) according to the panning round). Bound oligopeptide phage particles were eluted with 200 µL phage elution buffer per well for 30 min at 37 °C. The eluted phage titer was determined using 10 µL of the elution and a 10-fold dilution series.

2.2.5.5 Amplification of eluted phage

Eluted phage were amplified to use as input phage for the next panning round. The remaining 390 µL of the elution (see above) were used to infect 5 mL of an *E. coli* TOP10F' culture (OD_{600} = 0.5) for 30 min at 37 °C. Cells were plated on 15 cm 2x YT-GA agar plates and incubated at 37 °C overnight in order to allow amplification of eluted phage for the next panning round. The cells were scraped using 5 mL 2x YT-GA and used to inoculate a culture for phage amplification or was supplemented with 20 % (v/v) glycerol and stored at -80 °C. Phage amplification was performed in 30 mL scale as described above (2.2.3.1.2).

2.2.5.6 Screening of monoclonal oligopeptide phage

After the second and third panning round, respectively, 10-fold dilutions of eluted phage were used to infect 50 µL *E. coli* TOP10F' (OD_{600} = 0.5). The cells were plated on 2x YT-GA agar plates and incubated at 37°C overnight in order to obtain single colonies that allow screening of monoclonal oligopeptide phage. The monoclonal oligopeptide phage were produced in MTP scale (2.2.3.1.3) and analyzed for their reactivity with serum antibodies by monoclonal screening ELISA (2.2.7.2.1).

2.2.6 Recombinant protein expression

Recombinant protein expression was performed using *E. coli* BLR(DE), an *E. coli* strain bearing an IPTG inducible T7 polymerase gene. Expression of the recombinant proteins

was controlled by a T7 promoter using the expression vectors pET21a(+) for cytoplasmic and pET21a(+)-pelB for periplasmic expression.

Electro-competent *E. coli* BLR(DE3) were transformed with the respective expression vector as described above (2.2.2.2). As a starter culture, 30 mL of 2xYT-GA were inoculated with the transformed cells and incubated over night at 37 °C and 250 rpm (Infors HT). For protein production, 300 mL 2xYT-GA were inoculated (OD_{600} = 0.1) and cultivated at 37 °C and 250 rpm until logarithmic growth phase was reached (OD_{600} = 0.5). To induce protein expression, the cells were pelleted for 15 min at 4,000 xg (Sorval Centrifuge RC5B Plus, Rotor F9S), the supernatant was discarded and the cells were resuspended in 300 mL 2x YT-A supplemented with 1 mM IPTG. Recombinant protein expression followed for 6 h at 22 °C and 200 rpm (Infors HT). The cells were pelleted for 15 min at 4,000 xg and 4 °C (Sorval Centrifuge RC5B Plus, Rotor F9S) and the recombinant proteins were subsequently isolated form the periplasm or cytoplasm.

2.2.6.1 Periplasmic protein isolation

For periplasmic protein isolation, the cells were resuspended in 15 mL cold PE buffer and incubated for 20 min on ice thoroughly vortexing every 5 min. The cells were pelleted for 20 min at 4,000 xg (Sorval Centrifuge RC5B Plus, Rotor SLA-3000) and the periplasmic preparation (supernatant) was collected. The cells were resuspended in 15 mL cold OS buffer and incubated for 20 min on ice thoroughly vortexing every 5 min. The cells were pelleted for 20 min at 4,000 xg (Sorval Centrifuge RC5B Plus, Rotor SLA-3000) and the osmotic shock fraction (supernatant) was collected. If not used immediately for affinity chromatography, the fractions were stored at 4 °C.

2.2.6.2 Cytoplasmic protein isolation

For cytoplasmic protein isolation, the cells were resuspended in 15 mL cold IMAC binding buffer. Cells were disrupted using sonication (5x 30 s, 50 % intensity, MS72 sonotrode, Sonopuls 2200). The cell debris were pelleted for 20 min at 10,000 xg and 4 °C (Sorval Centrifuge RC5B Plus, Rotor SLA-3000) and the protein containing supernatant was transferred to a fresh tube and subject to affinity chromatography. If not used immediately the supernatant was stored at 4 °C.

2.2.6.3 Immobilized metal ion affinity chromatography (IMAC)

Recombinant proteins were purified using immobilized metal ion affinity chromatography (IMAC). Supernatant from cytoplasmic protein isolation was directly mixed with 1 mL Ni-sepharose and incubated for 45 min constantly mixing by inverting. The fraction from periplasmic protein isolation were combined and mixed with the same volume of IMAC binding buffer. Thirty mL were mixed each with 1 mL Ni-sepharose (Chelating Sepharose Fast Flow, GE Healthcare) and incubated for 45 min constantly mixing by inverting.

Sepharose with bound protein was pelleted for 10 min and 500 xg (Eppendorf centrifuge) and the supernatant was discarded. The sepharose was transferred and to flow column and washed with 25 mL IMAC binding buffer (supplemented with 10-50 mM imidazole). Proteins were eluted twice with 1 mL IMAC elution buffer each.

In order to remove EDTA from the protein elution, the proteins were dialyzed against PBS at 4 °C overnight (MWCO 3.5 - 14 kDa, according to protein MW).

2.2.6.4 Spectrophotometric determination of protein concentrations

Protein concentrations were determined using a spectrophotometer (NanoDrop ND1000, λ = 280 nm) applying the calculated molecular weight and molar extinction coefficient (according to Vector NTI Advance 11.5 or Geneious 4.8.5).

2.2.6.5 SDS-PAGE

The discontinuous SDS-PAGE was performed as outlined by Laemmli (Laemmli, 1970) using a stacking gel to focus the proteins with a subsequent separation in a separating gel. The composition of the stacking (Table 2-29) and separating gel (Table 2-30) is summarized in the following tables:

Table 2-29: Recipe for a stacking polyacrylamide gel

Component	Volume	
dH_2O	1.00	mL
Acrylamide Mix	0.26	mL
1 M Tris-HCl, pH 6.8	0.20	mL
SDS solution	15	µL
APS solution	15	µL
TEMED	2	µL

Table 2-30: Recipe for a separating polyacrylamide gel

Component	Volume 12 %	Volume 15 %	
dH_2O	1.3	0.9	mL
Acrylamide Mix	1.6	2.0	mL
1.5 M Tris-HCl, pH 8.8	1.0	1.0	mL
SDS solution	40	40	µL
APS solution	40	40	µL
TEMED	2	2	µL

The protein samples were mixed with SDS sample buffer (+ 2-mercaptoethanol) followed by incubation at 95°C for 10 min. The electrophoresis was performed at 200 V and 25 mA/gel for 45-60 min. After SDS-PAGE the gels were either stain with a Coomassie staining solution or blotted on a PVDF membrane.

2.2.6.6 Coomassie staining of polyacrylamide gels

For Coomassie staining, the polyacrylamide gels were incubated in Coomassie staining solution. The gels were microwave heated and incubated until the gel was completely stained. The staining solution was removed and the gel was washed with dH_20. The following destaining was performed by microwave heating of the gel in destaining solution and incubation with pieces of cotton tissue until the gel was destained und the protein bands were visible.

2.2.7 Immunological methods

2.2.7.1 Immunoblot of phage particles

The oligopeptide phage (4x 10^8 - 6x 10^9 cfu) proteins were separated on a 12 % polyacrylamide gel and blotted (20 V, 770 mA, 35 min) onto PVDF membranes. Membranes were blocked by incubating in MPBS at 4 °C overnight. Sera of fifteen 18 week-old $TNF^{\Delta ARE/+}$ and 19 C57BL/6N wt male mice were pooled. These pools and the reference sera used for selection (1:200 in MPBS containing 1.5 x 10^{10} cfu/mL M13K07 wildtype phage for competition) were incubated at 4 °C overnight. Membranes were stained with the sera for 2.5 h at room temperature and excess serum antibodies were removed by using washing buffer PBST (0.1 %). Bound serum antibodies were detected with a goat anti-mouse (IgA, IgG, IgM) antibody conjugated with HRP (antikoerper-online, ABIN376237, 1:4,000 in MPBS) for 1.5 h at room temperature. Immunoblots were developed using diaminobenzidine (DAB) substrate.

2.2.7.2 Enzyme-linked immunosorbent assay (ELISA)

ELISA was performed in 96-well (Costar MTP 96 Well Polystyrole, Corning) and 384-well (MTP 384 Well Polystyrole, Greiner Bio-One) scale using 100 µL/well and 30 µL/well total volume, respectively.

2.2.7.2.1 Monoclonal screening ELISA

In order to screen the clones after panning, monoclonal oligopeptide phage were produced as described above (2.2.3.1.2) and binding of serum antibodies was detected by ELISA.

To capture the oligopeptide phage in a screening setup with mouse serum, a rabbit anti-M13 antibody (pVIII specific, 1:5,000 in PBS, PA1-46334, Pierce Biotechnology) was immobilized on an ELISA plate at 4 °C overnight.

For a screening setup using human sera, a mouse anti-M13 antibody (pVIII specific, 1:1,000 in PBS, 61097, Progen) was immobilized at 4 °C overnight.

Wells were completely saturated with MPBST for 1 h at room temperature. The plates were washed 3 times with ELISA washing buffer before adding the monoclonal phage (40 µL in 100 µL MPBST). Phage particles were captured for 2 h at room temperature (50 µL precipitated phage in 100 µL MPBST). Remaining monoclonal phage were removed by 3 washing steps with ELISA washing buffer and the sera (1:500-1:1,000 in MPBST) were transferred to the wells containing immobilized oligopeptide phage particles and incubated for 2 h at room temperature. Excess serum antibodies were removed by 3 washing steps with ELISA washing buffer.

Bound murine serum antibodies were detected using a goat anti-mouse (IgA, IgG, IgM) antibody conjugated with horseradish peroxidase (HRP) (antikoerper-online, ABIN376237, 1:8,000 in MPBST) for 1.5 h at room temperature.

Bound human serum antibodies were detected using a goat anti-human (IgA, IgG, IgM) antibody conjugated with HRP (Dianova, 109-036-064, 1:5,000 in MPBST) for 1 h at room temperature.

Excess detection antibody was removed by 3 washing steps and the ELISA was developed with TMB substrate solution for up to 10 min. Signals were detected using an ELISA reader (450 nm, reference 620 nm).

2.2.7.2.2 ELISA based measurement of antibody response

2.2.7.2.2.1 96-well scale

The reactivity of mice sera with the selected oligopeptides was performed in 96-well format. Serum reactivity with the peptides was measured in doublets. The measured background binding to streptavidin was subtracted from the mean serum reactivity to compensate inter-individual differences in streptavidin background binding.

The 96-well ELISA plates were coated with 200 ng/well streptavidin at 4 °C overnight and the whole wells were blocked with BPBST. After three washing steps with ELISA washing buffer, 200 ng/well of synthetic biotinylated peptide was immobilized for 2 h at room temperature. Excess peptide was removed by three washing steps with ELISA washing buffer. The sera (1:100 dilution in MPBST) and incubated with the immobilized peptides and the streptavidin controls for 2.5 h at room temperature. Unbound serum antibodies were removed by three washing steps with ELISA washing buffer. The bound serum antibodies were detected with goat anti-mouse (IgA, IgG, IgM) antibodies conjugated with HRP (antikoerper-online, ABIN376237, 1:8,000 in MPBST) for 1.5 h at room temperature and the ELISA was developed with TMB substrate solution as described above and signals were detected using an ELISA reader (450 nm, reference 620 nm).

2.2.7.2.2.2 384-well scale

The reactivity of human sera with the selected oligopeptides and full-length proteins (oligopeptide phage or synthetic peptide) was performed in 384-well scale. Serum reactivity with the peptides was measured in triplicates. The measured mean background binding to M13K07ΔgIII Hyperphage (oligopeptide phage) or system background in blocked wells (synthetic peptides) was subtracted from the mean serum reactivity to compensate inter-individual differences in background binding.

For ELISA using oligopeptide phage, the 384-well ELISA plates were coated with mouse anti-M13 (pVIII specific) (1:1,000 in PBS, 61097, Progen) at 4 °C overnight and the wells were blocked with MPBST for 1 h at room temperature. After three washing steps with

ELISA washing buffer, 2.5x 10^7 cfu oligopeptide phage and M13K07ΔgIII Hyperphage was immobilized for 1.5 h at room temperature. Excess phage was removed by three washing steps with ELISA washing buffer.

For ELISA using recombinant protein and synthetic peptide, 30 ng/well (in PBS) were immobilized at 4 °C overnight. Control wells were filled with only PBS. The wells were blocked for 1 h at room temperature and washed 3 times using ELISA washing buffer.

The sera (1:500 dilution in MPBST) were transferred to the ELISA plates (Viaflo96, Integra) and incubated for 1.5 h at room temperature with the immobilized oligopeptide phage and proteins/peptides, respectively. Unbound serum antibodies were removed by three washing steps with ELISA washing buffer. The bound serum antibodies were detected with a goat anti-human (IgA, IgG, IgM) antibody conjugated with HRP (Dianova, 109-036-064, 1:5,000 in MPBST) for 1.5 h at room temperature and the ELISA was developed with TMB substrate solution as described above and signals were detected using an ELISA reader (450 nm, reference 620 nm).

3 Results

3.1 Metaproteome display in experimental ileitis (TNF$^{\Delta ARE/+}$ mouse model)

3.1.1 Construction of gut metaproteome phage display libraries

Three individual metaproteome phage display libraries were constructed from metagenomic DNA isolated from the caecum of three 8-week-old TNF$^{\Delta ARE/+}$ mice (the metagenomic DNA was a kind gift from Dr. Thomas Clavel, TU München, Germany) (Table 3-1). To increase starting material quantity, the metagenomic DNA was isothermally amplified by multiple displacement amplification (MDA) a suggested method for whole genome amplification (Taupp, Mewis, & Hallam, 2011). Whereas metagenomic DNA generally appears as a rather distinct band with approx. 5-10 kb when analyzed by agarose gelelectrophoresis, DNA after MDA resulted in branched DNA (Spits et al., 2006) strands that appeared as a smear in the range of 1 to 10 kb in an agarose gel (Figure 3-1).

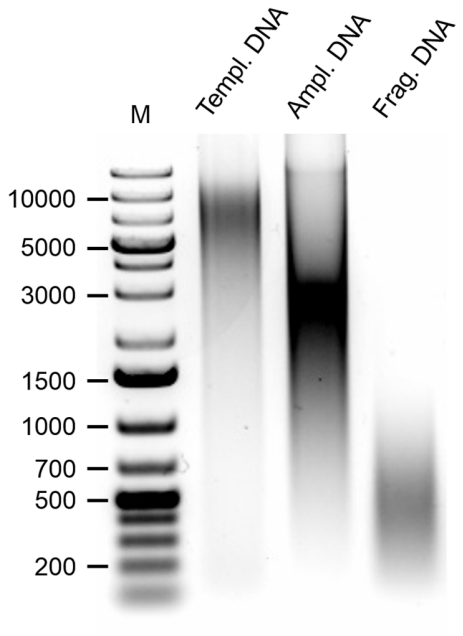


Figure 3-1: Agarose gel presenting the different levels of DNA preparation. (Templ. DNA = metagenomic template DNA for MDA, Ampl. DNA = MDA product, Frag. DNA = fragmented DNA by sonication)

The amplified gDNA was randomly fragmented by sonication yielding DNA fragments in the range of 100 bp - 1 kb (Figure 3-1). The DNA fragments were cloned into the phage display vector pHORF3 in a non-directional manner. Based on transformation rates, the

initial library diversities were 5.9x10^7, 1.9x10^7 and 2.5x10^7 independent clones with mean insert sizes of 335 bp, 284 bp and 354 bp, respectively. Insert rates were >85 % for all three libraries as analyzed by colony PCR (Figure 3-2).

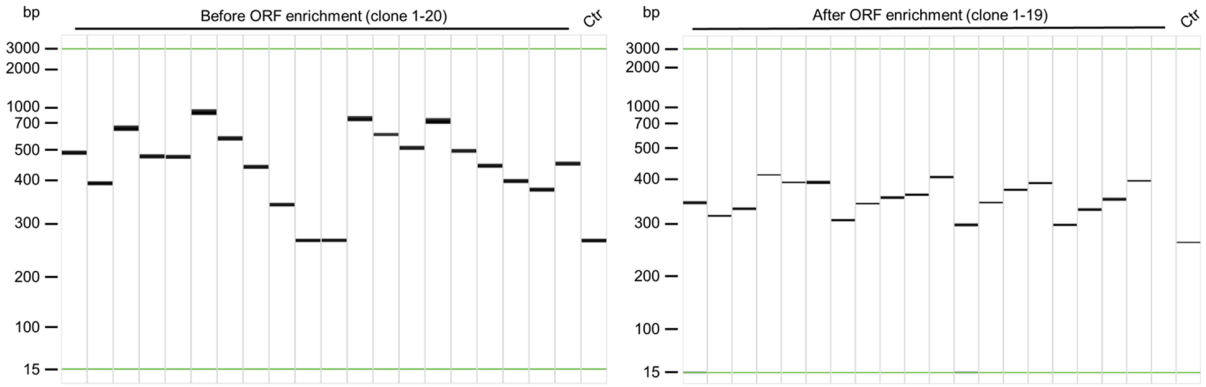


Figure 3-2: Representative PCR of single colonies before and after ORF enrichment. PCR products were analyzed on a Qiaxcel Advance capillary electrophoresis machine. (Ctr = empty library vector pHORF3).

Library quality was further enhanced by the enrichment of open reading frames (ORF) coupled to the display of the corresponding oligopeptides on phage particles. Inserts rates were 100% as analyzed by colony PCR, ORF enrichment was verified by Sanger DNA sequencing and ranged from 50 % to 80 %, the mean insert size was decreased after ORF enrichment with 92 bp – 117 bp (Figure 3-2).

Table 3-1: Summary of constructed metaproteome phage display libraries

Library		**1**	**2**	**3**
Approach		Mouse model		
Source		Caecum TNF$^{\Delta ARE/+}$	Caecum TNF$^{\Delta ARE/+}$	Caecum TNF$^{\Delta ARE/+}$
Initial library size (x10^7)		5.9	1.9	2.5
Before ORF enrichment	**Mean insert size (bp)**	335	284	354
	Insert rate (%)	90	85	85
After ORF enrichment	**Mean insert size (bp)**	92	102	117
	Insert rate (%)	100	100	100
	ORF rate (%)	75	50	80

3.1.2 NGS analysis of metaproteome phage display libraries

To obtain an overview of the starting oligonucleotide diversity prior to panning and reproducibility of library packaging, triplicate phage libraries of each of the three donor mice were sequenced. An average of 193,500 total reads and 122,000 unique reads were obtained per mouse per replicate (NGS and raw data processing was performed by Dr. Ilias Lagkouvardos, TU München, Germany). NGS analysis of the ORF enriched libraries revealed a size distribution ranging from 8 to 476 nt, with a mean sequence length of 65.8 ± 4.3 to 84.8 ± 1.9 nt (size limit due to sequencing specifications was approximately 500 nucleotides) Mean sequence length was comparable to the mean sequence length obtained from a small sample (19 – 20 clones) by Sanger sequencing and represents a decrease in insert sizes compared to the initial libraries before ORF enrichment.

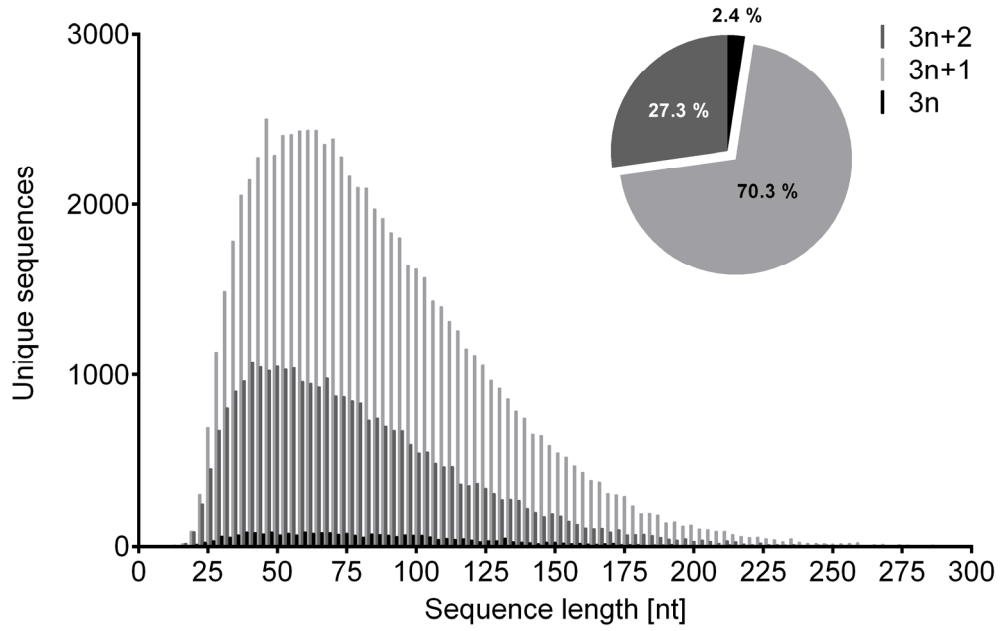


Figure 3-3: Size distribution histogram of ORF enriched oligopeptide phage library determined by NGS analysis. Sequences with 3n+1 nucleotides (nt) were enriched over sequences with 3n+2 and 3n nt. The histogram was trimmed at 300 nt but sequences with up to 476 nt were detected.

Random fragmentation of DNA for library construction results in DNA fragments with sizes of 3n, 3n+1 and 3n+2 nucleotides. Sequences with a length of 3n+1 were enriched over sequences with 3n+2 and 3n nucleotides (70.3 % vs. 27.3 % vs. 2.4 % of unique reads) (Figure 3-3). Due to the vector design (Suppl. Figure 1), only inserts with 3n+1 nucleotides can result in ORFs with phage pIII. Hence, NGS data strongly supported efficacy of the

ORF enrichment and was comparable to the ORF rate obtained from Sanger sequencing of a small sample of randomly analyzed clones (19 – 20 clones).

Approximately 30 % of the reads (54,000 of 176,000, represented on average by 30,000 peptides) corresponded to redundant peptides (i.e., those with more than one copy; median copy number was 2, showing that most of the redundant peptides had low copy numbers; maximum copy number ranged between 44 and 163, depending on the library). The 20 most abundant clones represented less than 0.5 % of total reads.

Across different library preparations from the same sample, only 8-10 % of the sequences were shared between all three replicates and another 6-8 % between each of the replicate pairs (overlap between duplicates was 18-23 %) (Figure 3-4) (data was provided by Dr. Ilias Lagkouvardos and Dr. Thomas Clavel, TU München, Germany).

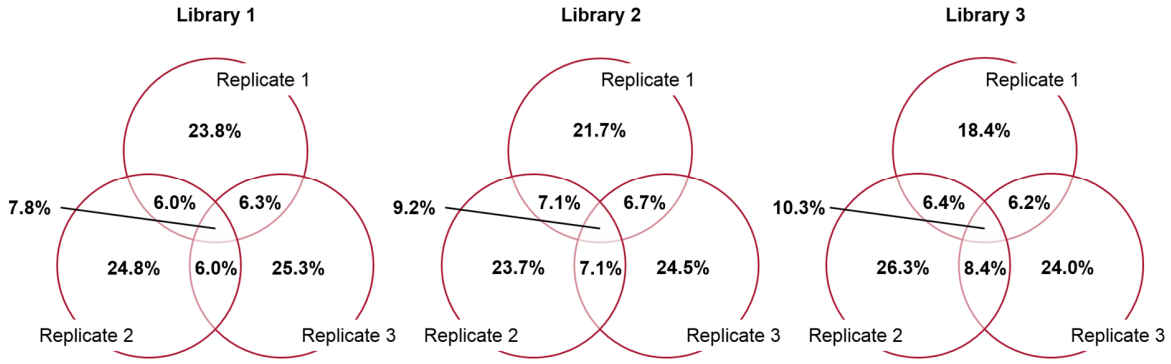


Figure 3-4: Venn's diagram showing the percentage of oligonucleotides that are shared between or unique to the ORFeome phage display library replicates. The percentages refer to a total number of approximately 265,000 oligonucleotide reads. (Data was provided by Dr. Ilias Lagkouvardos and Dr. Thomas Clavel, TU München, Germany.)

These findings suggest that the total diversity of oligonucleotides existing in the libraries was not covered by either the individual library preparation or more likely by the sequencing depth, and it exceeded 265,000 detected sequences per library. Therefore, a final statement about reproducibility of library packaging is not possible. But these data show that the constructed library displayed a high degree of diversity (unique peptides dominated the libraries), which demonstrates their suitability as starting materials for peptide selection.

3.1.3 Selection of immunogenic oligopeptides from metaproteome phage display libraries

For the selection of microbiota-derived immunogenic oligopeptides from metaproteome phage display libraries, three independent pannings, one for each mouse, using immobilized serum antibodies (IgA, IgG and IgM in combination) were performed. Matched samples (i.e. one individual library constructed from a given mouse gut content and the corresponding serum sample) were used. After three consecutive rounds of panning, 92 clones from each panning were screened by monoclonal phage ELISA for an antibody response of the designated serum specific towards the displayed oligopeptide (data not shown). In total, 16 individual clones reactive with the designated mouse serum were selected (Table 3-2).

Some selected clones were found in multiple copies among the 276 screened clones. In particular, clone JOZ156-A6 was dominantly enriched during the panning procedure, but was later found to be a universal immunoglobulin binding protein (data not shown). The other selected fragments ranged from 31 to 328 bp in size and originated from unknown sequences. Blastn did not return any annotation for the selected clones except from clone JOZ158-E11 which was found to have 84 % identity with an annotated sequence from *Roseburia hominis*. Protein homologies (blastp) were only found for clones JOZ157-E1 and JOZ158-E11. Clone JOZ156-A6, JOZ156-A8, JOZ156-F7, and JOZ157-C10 were annotated as predicted proteins from *Blautia* species.

3.1.4 Validation of oligopeptide immunogenicity by immunoblot

Blotted proteins of the selected oligopeptide phage particles were incubated with pools of each 15 $TNF^{\Delta ARE/+}$ and 19 wt C57BL/6N mouse sera at the age of 18 weeks. All $TNF^{\Delta ARE/+}$ mice were inflamed in the distal ileum and proximal colon, as confirmed by histological scoring (3.9 ± 1.3 vs. 0.03 ± 0.08 (distal ileum) and 4.3 ± 1.7 vs. 0.05 ± 0.13 (proximal colon)) (histological scoring was performed by Dr. Sigrid Kisling, TU München, Germany).

Table 3-2: Summary of selected clones

Clone	Insert size (bp)	Redundancy	Origin (blastn)	Protein homology (blastp)	GenBank Accession
JOZ156-A3	43	1	Unknown	-	-
JOZ156-A6	298	32	Unknown	Predicted protein (*Blautia* sp.) (E-value 7e-24, 66 % identity)	n.d.
JOZ156-A8	319	1	Unknown	Predicted protein (*Blautia* sp.) (E-value 3e-27, 57 % identity)	n.d.
JOZ156-B11	58	6	Unknown	-	-
JOZ156-F7	328	1	Unknown	Predicted protein (*Blautia* sp.) (E-value 3e-29, 58 % identity)	n.d.
JOZ156-G7	85	2	Unknown	-	-
JOZ156-H5	58	1	Unknown	-	-
JOZ157-C10	211	3	Unknown	Predicted protein (*Blautia* sp.) (E-value 7e-18, 72 % identity)	n.d.
JOZ157-E1	202	1	Unknown	ABC transporter permease (*Eubacterium* sp. ER2) (E-value 2e-21, 62 % identity)	WP_033126529
JOZ157-G6	58	1	Unknown	-	-
JOZ158-C11	76	1	Unknown	-	-
JOZ158-C12	46	1	Unknown	-	-
JOZ158-D10	31	2	Unknown	-	-
JOZ158-E11	121	2	*Roseburia hominis* (E-value 3e-9, 84 % identity)	Phosphoribosylaminoimidazole-succinocarboxamide synthase (*Blautia* sp.) (E-value 4e-20, 100 % identity)	CDA05871.1
JOZ158-G8	163	1	Unknown	-	-
JOZ158-H8	52	1	Unknown	-	-

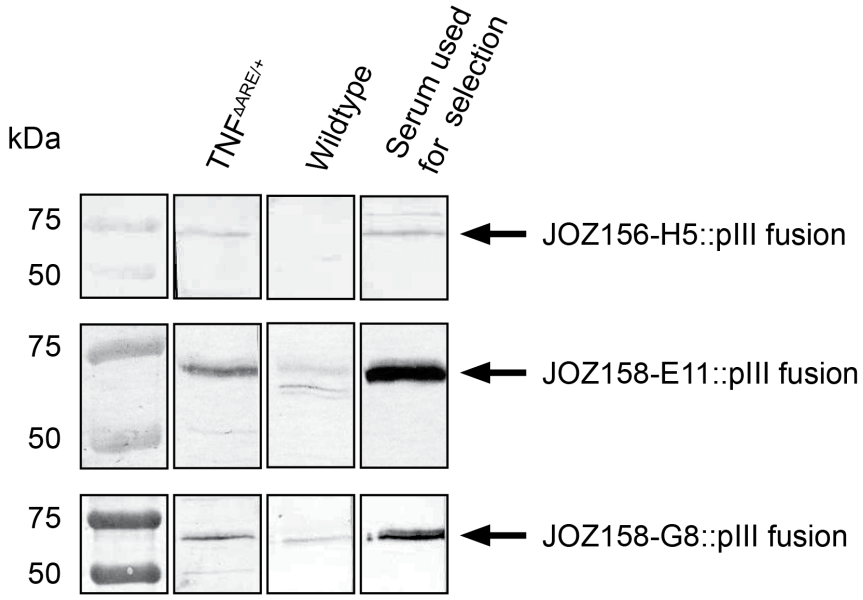


Figure 3-5: Immunoblot of oligopeptide phage particles stained with mouse serum. Oligopeptide:pIII fusion protein was stained with serum pools (1:200 dilution) from TNF$^{\Delta ARE/+}$ and wildtype mice. Bound antibodies were detected with an anti-mouse (IgA, IgG, IgM) secondary antibody.

Detection of bound serum antibodies resulted in a stronger staining of the oligopeptide::pIII fusion protein for three of the 16 candidate oligopeptide phage clones (Figure 3-5). These peptides (JOZ156-H5, JOZ158-E11, and JOZ158-G8) were considered to be the most promising candidates for detection of specific circulating antibodies in TNF$^{\Delta ARE/+}$ mice and were synthesized as biotinylated peptides.

3.1.5 Validation of TNF$^{\Delta ARE/+}$ specific immune responses by ELISA

3.1.5.1 Conventional housing

Even though immunogenicity was validated using oligopeptide phage and serum pools, further characterization was performed using synthetic peptides and individual sera of the TNF$^{\Delta ARE/+}$ and control cohorts. Peptides were immobilized via streptavidin coupling and incubated with the individual sera. Bound anti-peptide serum antibodies were detected by ELISA using a secondary anti-mouse antibody. Some sera of mostly the TNF$^{\Delta ARE/+}$ cohort reacted with streptavidin alone and competition with streptavidin in solution did not decrease the background binding of streptavidin. Therefore, a background subtraction step was implemented to exclude false positive reactive sera. The measured antibody

responses against all three peptides were significantly higher in TNF$^{\Delta ARE/+}$ mice compared to the wildtype, despite marked inter-individual differences (Figure 3-6).

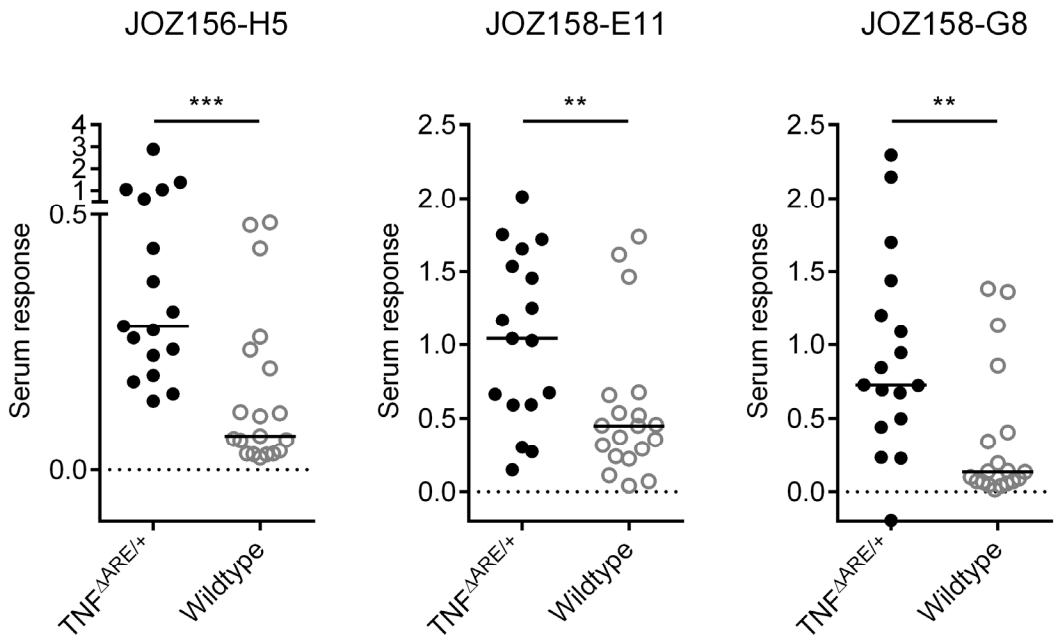


Figure 3-6: Specific reactivity of TNF$^{\Delta ARE/+}$ cohort from conventional housing with the selected biomarker candidates. Serum response was determined by ELISA and plotted as mean of duplicate measurement. A background subtraction of streptavidin binding was included. The median cohort reactivity is indicated by a black line. Peptides were immobilized by streptavidin and bound serum antibodies from single sera (1:100 dilution) were detected with an anti-mouse (IgA, IgG, IgM) antibody. (Non-parametric Mann-Whitney test; ** = p < 0.01; *** = p < 0.001)

Compared to peptide JOZ156-H5, antibody responses against the peptides JOZ158-E11 and JOZ158-G8 were generally stronger and reactivity of some control sera was increased. Nevertheless, the serum reactivity of the TNF$^{\Delta ARE/+}$ cohort against both peptides (JOZ158-E11 and JOZ158-G8) was significantly stronger compared to the control sera ($p < 0.01$). Consequently, all three isolated peptides were considered as biomarker candidates for the TNF$^{\Delta ARE/+}$ mouse model. Peptide JOZ158-E11 was identified as a peptide from phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR synthase) of a *Blautia* species. At the DNA level however (Table 3-2), there was only similarity to the DNA sequence of the SAICAR synthase gene from *Roseburia hominis*. The peptides JOZ156-H5 and JOZ158-G5 returned no annotation by using both blastn and blastp.

3.1.5.2 Specific pathogen free housing (SPF) and human IBD samples

In order to assess the influence of the gut microbiome on the development of specific antibody responses against the selected biomarker candidates, serum response was measured by ELISA using samples from TNF$^{\Delta ARE/+}$ and matched wildtype mice kept under specific pathogen free (SPF) housing conditions. Ileal inflammation in mice kept under SPF conditions was generally weaker than in mice kept under conventional housing conditions as determined by histological scoring (3.0 ± 2.3 vs. 3.9 ± 1.3) and was absent in the proximal colon (0.1 ± 0.3 vs. 4.3 ± 1.7) (histological scoring was performed by Dr. Sigrid Kisling, TU München, Germany). Weaker gut inflammation in mice from SPF housing compared to conventional housing was reflected by overall weaker serum response. Antibody titers against the oligopeptides JOZ156-H5 and JOZ158-E11 were increased in TNF$^{\Delta ARE/+}$ mice when compared to wildtype controls but was very weak for JOZ156-H5 and absent for JOZ158-G8 (Figure 3-7).

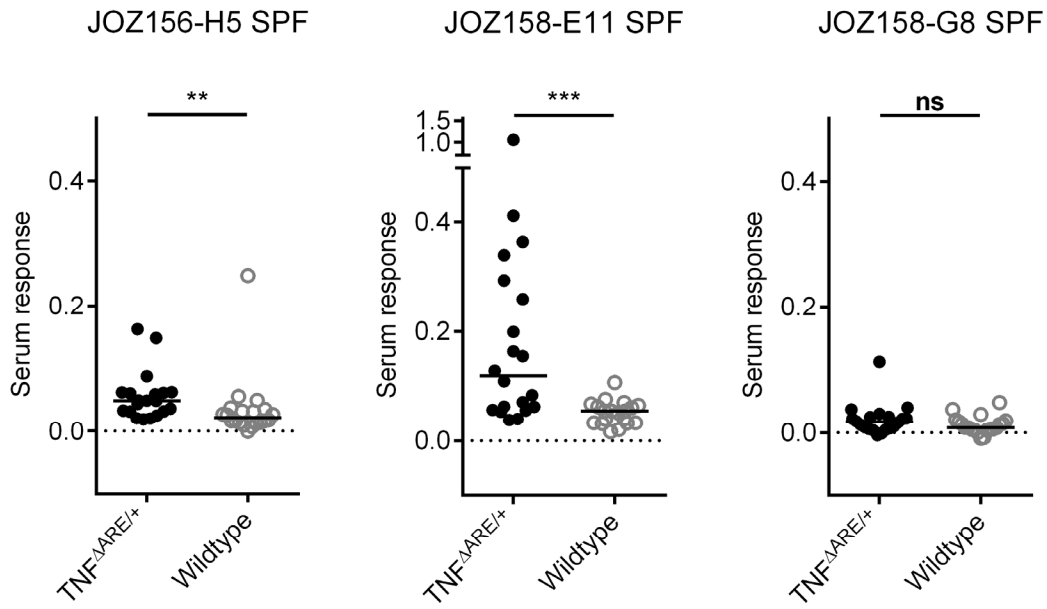


Figure 3-7: Reactivity of TNF$^{\Delta ARE/+}$ cohort from SPF with the selected biomarker candidates. Serum response was determined by ELISA and plotted as mean of duplicate measurement. A background subtraction of streptavidin binding was included. The median cohort reactivity is indicated by a black line. Peptides were immobilized by streptavidin and bound serum antibodies from single sera (1:100 dilution) were detected with an anti-mouse (IgA, IgG, IgM) antibody. (Non-parametric Mann-Whitney test; ** = p < 0.01; *** = p < 0.001)

The relatively low but specific antibody responses against two out of the three oligopeptides may indicate a correlation between intestinal inflammation and antibody

response. The lack of specific antibody titers against the peptide JOZ158-G8 suggests that the peptide originates from a bacterial species that was not present in SPF housing.

3.1.6 Time dependent occurrence of biomarker candidates

TNF$^{\Delta ARE/+}$ mice are known to develop intestinal inflammation gradually from the age of 4-6 weeks on with plateauing around 12 weeks of age (Baur et al., 2011; Kontoyiannis et al., 1999). An overall weaker serum response against two out of the three biomarker candidates in mice kept under SPF housing conditions and less intestinal inflammation suggested a correlation between inflammatory state and specific antibody responses. In order to test whether the systemic antibody response towards the three identified gut microbiome-derived peptides gradually developed over age and inflammatory state, serum reactivity was monitored using ten mice of each genotype sacrificed at week 4, 6, 8, 12, and 18 of age. Intestinal inflammation was assessed by histopathological scoring (histological scoring was performed by Dr. Sigrid Kisling, TU München, Germany) and anti-JOZ156-H5, anti-JOZ158-E11, and anti-JOZ158-G8 serum antibody titers were measured by ELISA (Figure 3-8).

Intestinal inflammation was observed by week 6 and the inflammatory state increased with older age of the TNF$^{\Delta ARE/+}$ mice, whereas no inflammation was observed in wildtype controls (Figure 3-8A). Inter-individual differences in serum response were observed for all three analyzed oligopeptides. Serum levels of anti-JOZ156-H5 antibodies were increased in 4-, 6-, 8-, 12- and 18-week-old TNF$^{\Delta ARE/+}$ mice when compared to the matched wildtype controls and tended to gradually develop increased antibody titers with age and inflammation (Figure 3-8B). Levels of anti-JOZ158-E11 antibodies were relatively high in 8-week-old TNF$^{\Delta ARE/+}$ mice, but were not further increased thereafter and were not abundant in the serum of 18-week-old mice (Figure 3-8C). Anti-JOZ158-G8 antibodies were only present in a fraction of TNF$^{\Delta ARE/+}$ samples and there was only a trend towards increased response from week 8 on (Figure 3-8D). There was no clear evidence of a direct correlation of increased intestinal inflammation and serum response against the selected biomarker candidates although there was a trend of increased serum response with increased intestinal inflammation. Overall serum response in the measured cohorts was very heterogeneous and specific antibody titers were not measured in the same individuals over time but in different individuals at different time points.

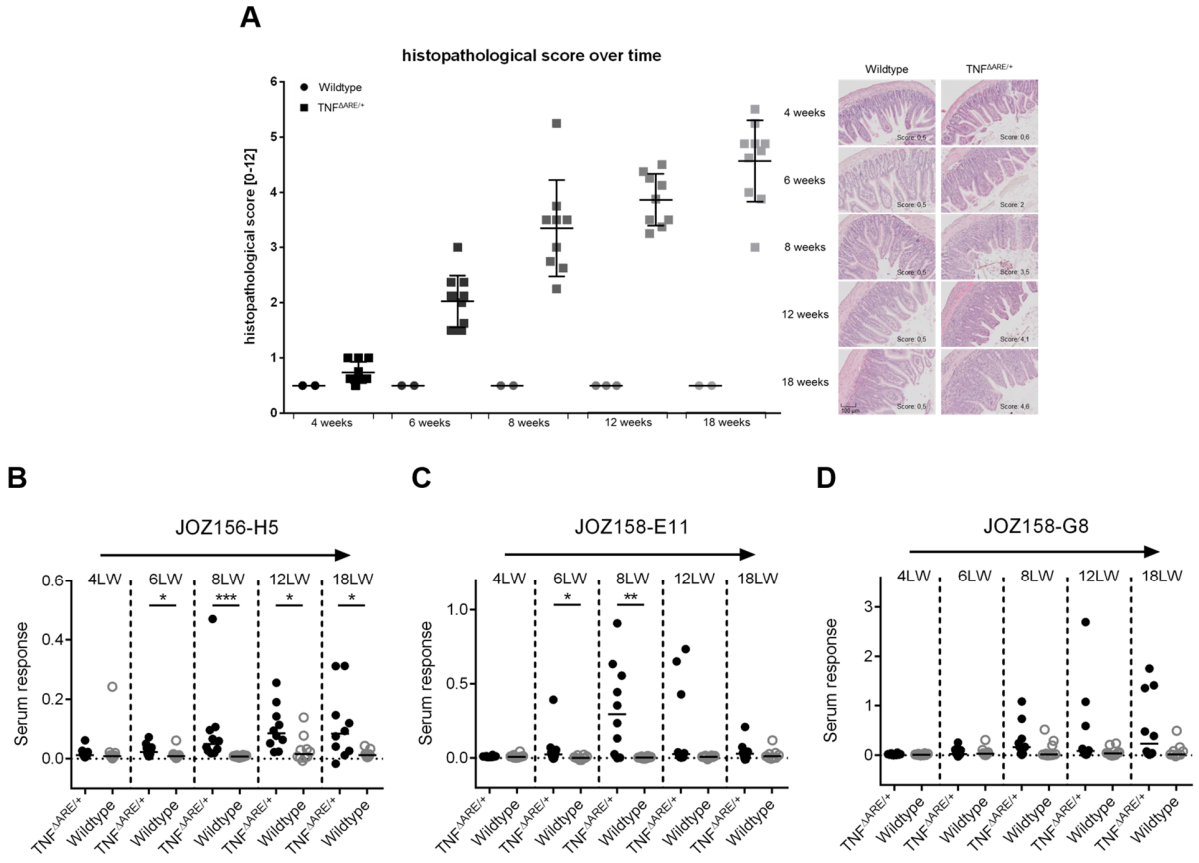


Figure 3-8: Time-course of ileal inflammation and serum response against the selected biomarker candidates. Serum response of mice sacrificed at life week 4, 6, 8, 12 and 18 was determined by ELISA and plotted as mean of duplicate measurement. A background subtraction of streptavidin binding was included. The median cohort reactivity is indicated by a black line. Peptides were immobilized by streptavidin and bound serum antibodies from single sera (1:100 dilution) were detected with an anti-mouse (IgA, IgG, IgM) antibody. (Non-parametric Mann-Whitney test; ** = p < 0.01; *** = p < 0.001) (Figure presenting histopathological scores and H&E staining was provided by Sarah Just, TU München, Germany)

Nevertheless, the results from the TNF$^{\Delta ARE/+}$ mouse model of experimental ileitis represented a proof-of-concept for the selection of gut microbiome-derived oligopeptides that have induced specific antibody responses in mice suffering from intestinal inflammation.

3.2 Metaproteome display in inflammatory bowel disease (clinical IBD samples)

For the selection of potential novel biomarker candidates for IBD, the ORFeome phage display technology was applied on human IBD samples following two different strategies. The selections were performed using an unmatched and a matched approach. As unmatched approach, metaproteome phage display libraries were construction from gut metagenomic DNA of two CD patients and oligopeptides were selected using serum pools

of different IBD patients based on disease phenotype. The IBD serum donors were in disease remission to only select biomarker candidates that are IBD specific and also detectable in remission state. For the matched approach, metaproteome phage display libraries were constructed from the gut metagenomic DNA of each two CD and UC patients and serum antibodies from the same individuals were used for the selection of candidate oligopeptides.

3.2.1 Construction of gut metaproteome phage display libraries

The construction of the metaproteome phage display libraries from gut metagenomic DNA of IBD patients was performed analogue to the construction of the libraries from the gut metagenomic DNA of $TNF^{\Delta ARE/+}$ mice. In total, seven metaproteome phage display libraries were constructed (Table 3-3). Three libraries were constructed for an unmatched panning approach (library 1-3) and four libraries were constructed for a matched panning approach (library 4-7). (The metagenomic DNA was a kind gift from Dr. Patricia Lepage, INRA, Joey-en-Josas (library 1-3) and Dr. Harry Sokol, St. Antoine Hospital, Paris (library 4-7)).

Metagenomic approaches demand for large libraries and in order to further increase the library sizes, the construction of library 4-7 was performed using the *E. coli* SS320 strain allowing higher transformation rates compared to the previously used *E. coli* TOP10F'. The *E. coli* SS320 strain was previously proven to be applicable to the ORFeome phage display technology and allowed transformations rates of more than 10^8 cfu/transformation following the same procedure as described in the present work (Felix Pritsch, Master thesis, TU Braunschweig, 2015).

Based on transformation rates, the initial library diversities for the unmatched panning approach (1-3) were 6.3x 10^7, 8.6x 10^7 and 5.7x 10^7 independent clones with mean insert sizes of 300 bp, 370 bp and 350 bp, respectively. Although the construction of library 4-7 was performed using an *E. coli* strain with higher competence, initial library sizes were 5.3x 10^7, 1.8x 10^7, 3.7x 10^7 and 3.2x 10^7 with mean insert sizes between 285 bp and 405 bp and were not increased compared to library 1-3. Notably, insert rates of library 4-7 were also lower (41 – 75 vs. 90 – 95 %) suggesting a correlation between insert rate and transformation rate which was also observed before (data not shown).

Table 3-3: Summary of constructed metaproteome phage display library from human gut content

Library		1	2	3	4	5	6	7
Approach		unmatched			matched			
Source		Colonic lumen CD	Colonic lumen CD	Caecal lumen CD	Stool CD	Stool CD	Stool UC	Stool UC
Initial library size (x10^7)		6.3	8.6	5.7	5.3	1.8	3.7	3.2
Before ORF enrichment	Mean insert size (bp)	300	370	350	285	300	365	405
	Insert rate (%)	90	95	95	75	41	55	71
After ORF enrichment	Mean insert size (bp)	115	115	-	190	170	205	230
	Insert rate (%)	100	100	-	95	100	100	100
	ORF rate (%)	61	50	-	55	70	52	61

ORF enrichment and production of phage particles resulted in very low phage titers for library 3 (at least more than two orders of magnitude lower). DNA sequencing of random clones of the initial libraries 1-3 and BLAST analysis revealed that library 3 contained >85% human DNA sequences whereas the fraction of human DNA sequences in library 1 and 2 was 28% and 40% (n = 30), respectively. No human-derived DNA sequences were detected in initial libraries 4-7. High content of human DNA sequences in the initial libraries may hamper the packaging of oligopeptide phage (accompanied by ORF enrichment) due to more intergenic DNA content of eukaryotic DNA sequences. ORF enrichment of the remaining metaproteome phage display libraries was verified by Sanger sequencing and ranged from 50 – 61 % to 52 – 70 % (library 4-7) with insert rates of mostly 100 % demonstrating the successful elimination of empty phagemid constructs from the metaproteome phage display libraries.

The library sizes and quality (insert and ORF rate) of ORF enriched metaproteome phage display libraries was comparable to the libraries from the proof-of-concept using the $TNF^{\Delta ARE/+}$ mouse model and were subsequently used for the selection of candidate oligopeptides using patient sera.

3.2.2 Selection of immunogenic oligopeptides from metaproteome phage display libraries

3.2.2.1 Unmatched samples

For the selection of candidate biomarker oligopeptides, 25 sera from CD patients and 25 sera from UC patients were grouped based on localization of intestinal inflammation and smoking habits (4 groups from CD patients and 2 from UC patients). (The patient sera were a kind gift from Dr. Harry Sokol, St. Antoine Hospital, Paris, France). The grouped sera were pooled. Using the immobilized antibodies (IgA, IgG and IgM in combination) from these serum pools, six individual pannings were performed. Panning 1-4 was performed using sera from CD patients and panning 5-6 was performed with UC patient sera. After three consecutive rounds of panning, 92 clones from each panning were screened by monoclonal phage ELISA for a specific antibody response of the designated serum pool (data not shown). In total, only 6 individual clones reactive with the designated serum pool were selected and subsequently identified by DNA sequencing (Table 3-4).

Table 3-4: Summary of selected clones applying an unmatched panning approach.

Clone	Panning	Insert size (bp)	Redundancy	Origin (blastn)	Protein	GenBank Accession
JOZ191-E5	3	73	1	*Escherichia coli* (E value 1e-13)	-	-
JOZ191-F12	3	109	1	*Bacteroides vulgatus* (E value 3e-49)	Putative LPS biosynthesis related UDP-galactopyranose mutase	ALK83393.1
JOZ191-G8	3	100	1	*Escherichia coli* (E value 1e-43)	Coproporphyrinogen III oxidase	ALY14458.1
JOZ192-A10	4	64	1	*Salmonella enterica* (E value 4e-24)	Guanylate kinase	AKF91369.1
JOZ193-A9	5	55	2	*Bacteroides dorei* (E value 4e-19)	-	-
JOZ194-C4	6	103	1	*Escherichia coli* (E value 7e-46)	nicotinate phosphoribosyl transferase	AIZ86460.1

Panning 1 and 2 (CD patient sera) did not reveal any oligopeptides that were specifically reactive with the designated serum pools. The coding DNA sequences of the selected oligopeptides from panning 3-4 (CD patient sera) and panning 5-6 (UC patient sera) ranged from 55 – 109 bp in size and only the oligopeptide JOZ193-A9 was found in another copy. A nucleotide BLAST analysis of the selected sequences returned annotations from *Bacteroides* species, *E. coli* and *Salmonella* species and except from JOZ191-E5 that seemed to be derived from a non-coding plasmid DNA sequence and JOZ193-A9 that did not translate in the correct frame to match a protein annotation, all selected sequences were linked with a protein annotation in GenBank.

In order to identify the most promising biomarker candidates, the reactivity of the different serum pools with all selected oligopeptides was determined by ELISA, reactivity was scored from 0 - 3 and presented as heat map (Table 3-5). Based on the reactivity of the different serum pools, the clones JOZ191-F12 and JOZ194-C4 were excluded from the panel for further characterization with single IBD sera.

Table 3-5: Heat map of serum reactivity of different IBD serum pools with the selected biomarker candidates. Clones marked with asterisk were further analyzed.

	Pool 1 (CD)	Pool 2 (CD)	Pool 3 (CD)	Pool 4 (CD)	Pool 5 (UC)	Pool 6 (UC)	Control pool (health)
JOZ191-E5*	0	0	2	0	0	0	0
JOZ191-F12	1	0	0	0	0	1	0
JOZ191-G8*	1	0	3	1	0	0	0
JOZ192-A10*	1	0	2	2	0	0	0
JOZ193-A9*	2	0	0	1	3	3	1
JOZ194-C4	1	0	0	0	0	1	0
Hyperphage control	0	0	0	0	0	0	0

3.2.2.2 Matched samples

In order to select gut microbiota derived immunogenic oligopeptides in a matched panning approach, the ORF enriched libraries were pooled according to their disease phenotype (CD, library 4 and 5/UC, library 6 and 7) and four individual pannings were performed

using immobilized serum antibodies (IgA, IgG and IgM in combination) from the IBD patients that served as donors for the gut metagenomic DNA.

Previous pannings resulted in enrichment of also universal immunoglobulin binding proteins (data not shown) and phage titers suggested a large enrichment of specifically bound oligopeptides already in panning round two (Suppl. Figure 2). Therefore, the metaproteome phage display libraries were precleared from universal immunoglobulin binding oligopeptides and peptides that were not specific for IBD by incubation with immobilized serum antibodies from healthy donors. To keep a broader diversity of selected oligopeptides, 92 clones from each panning were already screened after the second panning round by monoclonal phage ELISA (data not shown). Antibody responses were measured with the designated patient serum used for panning and a serum pool of healthy donors to exclude universal immunoglobulin binding proteins and oligopeptides reactive with sera from healthy individuals.

In total, 32 individual clones were selected including two universal immunoglobulin binding proteins (complete list in Suppl. Table 1). In order to exclude oligopeptides that were only reactive with the single patient serum used for the selection, cross-reactivity with the 4 patient sera used for panning and 6 different IBD serum pools as well as a healthy control pool was determined by ELISA. Nine different clones seemed to be specific for IBD (Table 3-6). However, clone JOZ241-A4 and JOZ243-E9 were identical overlapping clones with JOZ241-A4 having 8 additional N-terminal amino acids and JOZ243-E9 one additional C-terminal amino acid. The selected specific sequences ranged from 31 bp to 349 bp. Clone JOZ243-F9 was not in-frame with phage pIII and therefore should not result in functional phage particles. However, the sequence of JOZ243-F9 contained another start codon downstream of the pelB leader and translation from this start codon resulted in an ORF with phage pIII that also contained an N-terminal signal peptide recognized by *E. coli* as predicted by SignalP 4.1 analysis (Petersen, Brunak, von Heijne, & Nielsen, 2011). The selection of out-of-frame sequences that contained alternative start codons resulting in ORFs with phage pIII and endogenous N-terminal signal peptides was described before (Connor et al., 2016).

Table 3-6: Selection of candidate oligopeptides with potential IBD specific features

Clone	Panning	Insert size (bp)	Copies	Origin (blastn)	Protein	GenBank Accession
JOZ240-D7	CD 1	181	1	*Escherichia coli* (E value 3e-88)	outer membrane lipoprotein	AKK49496.1
JOZ240-H2	CD 1	121	1	*Bifidobacterium longum* subsp. *Longum* (E value 3e-55)	-	-
JOZ241-A4	CD 2	229	1	*Bifidobacterium adolescentis* (E value 1e-111)	glycine-rich cell wall protein	AJE06446.1
JOZ241-C4	CD 2	349	1	*Enterococcus faecalis* (E value 2e-176)	chlorohydrolase family protein	AAO81020.1
JOZ242-C9	UC 1	328	1	*Bifidobacterium breve* (E value 7e-170)	Hypothetical protein	ALE13104.1
JOZ242-G4	UC 1	217	1	Unknown	-	-
JOZ243-F12	UC 2	31	3	Unknown	-	-
JOZ243-E9	UC 2	208	1	*Bifidobacterium adolescentis* (E value 5e-100)	glycine-rich cell wall protein	AJE06446.1
JOZ243-F9	UC 2	296	1	*Bifidobacterium adolescentis* (E value 9e-149)	1,4-beta-N-acetylmuramidase	AII76920.1

Nucleotide BLAST analysis of the selected sequences did not return any annotation for the clones JOZ242-G4 and JOZ243-F12 but the remaining sequences were found to originate from *E. coli*, *Enterococcus faecalis* and *Bifidobacterium* species.

3.2.3 Validation of IBD specific immune responses by ELISA

In order to assess the specificity of the selected oligopeptides, the reactivity of 25 sera from CD patients, 25 UC patient sera and 9 healthy control sera was measured by ELISA. Reactivity of the oligopeptides of the unmatched panning approach was determined using synthetic peptides whereas specific reactivity against oligopeptides from the matched panning approach was determined based on oligopeptide phage.

3.2.3.1 Unmatched samples (synthetic peptides)

No oligopeptide showed strong serum reactivity specific for one of the IBD phenotypes or IBD in general (Figure 3-6).

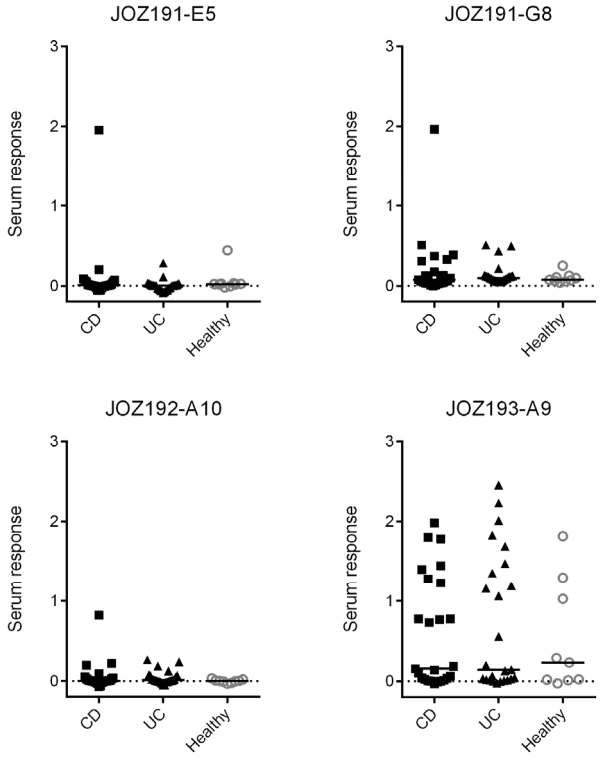


Figure 3-9: Reactivity of IBD serum cohorts with biomarker candidates selected by an unmatched panning approach (synthetic peptides). Serum response was determined by ELISA and plotted as mean of duplicate measurement. A background subtraction of streptavidin binding was included. The median cohort reactivity is indicated by a black line. Peptides were immobilized by streptavidin and bound serum antibodies from single sera (1:500 dilution) were detected with an anti-human (IgA, IgG, IgM) antibody.

Notably, the IBD cohorts contained at least one serum that reacted very strongly with the respective oligopeptide whereas this sample was part of the pool used for selection. The selection of the respective oligopeptides was presumably only due to the very high antibody titers in one individual serum of the pool. At least, some IBD sera also reacted with the peptide JOZ191-G8 even though it was not significant when compared to the healthy control cohort. Applying an unmatched panning approach did not result in the selection of oligopeptides that were highly sensitive to the IBD serum cohorts.

3.2.3.2 Matched samples (oligopeptide phage)

Selecting immunogenic oligopeptides using serum antibodies and the metaproteome phage display library from the same donor did not result in candidate biomarkers that reacted with the majority of samples in the IBD serum cohorts (Figure 3-10).

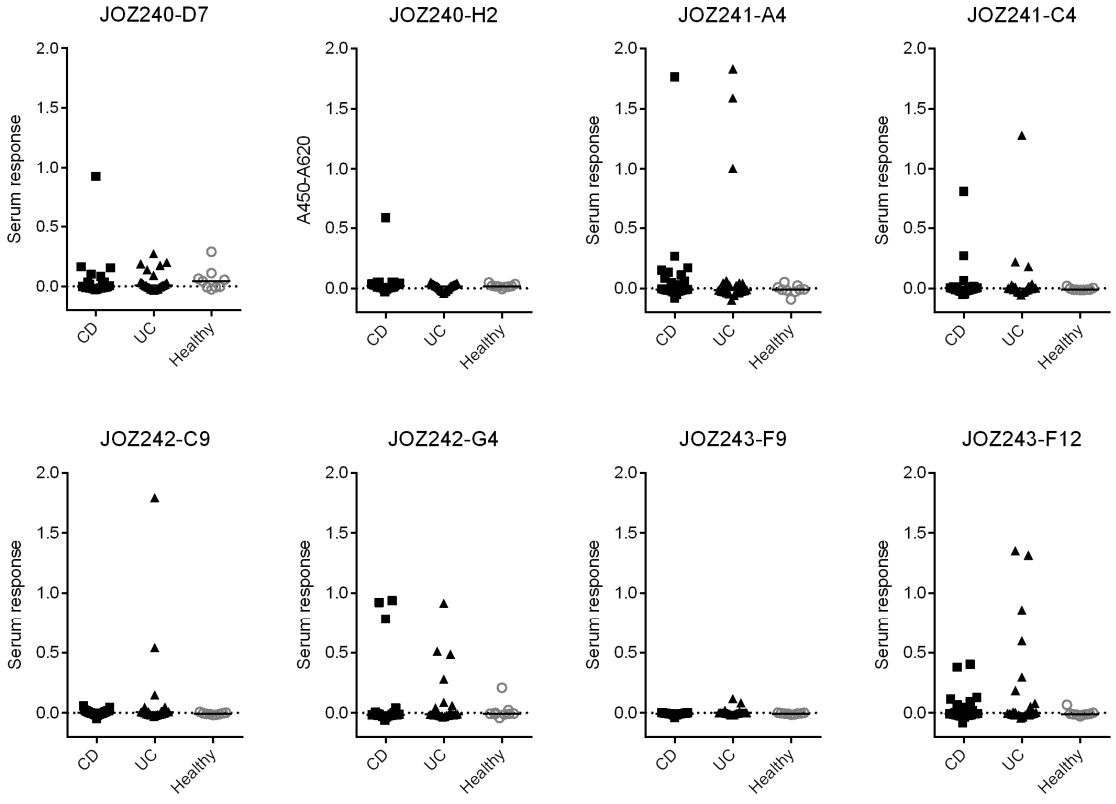


Figure 3-10: Reactivity of IBD serum cohorts with biomarker candidates selected by a matched panning approach (oligopeptide phage). Serum response was determined by ELISA and plotted as mean of triplicate measurement. A background subtraction of phage binding was included. The median cohort reactivity is indicated by a black line. Oligopeptide phage (2.5 x 10^7 cfu) were immobilized by an anti-M13 antibody and bound serum antibodies from single sera (1:500 dilution) were detected with an anti-human (IgA, IgG, IgM) antibody.

The cohorts contained the sera that were used for the selection and these single sera reacted strongly with the respective oligopeptide phage. Only a few other sera of the IBD cohorts were also reactive with some of the selected candidate biomarkers. But there was no biomarker candidate that suggested promising features for diagnostic performance in IBD.

However, except from oligopeptide JOZ240-D7, sera from the healthy control cohort did not react with the candidate biomarkers at all, i.e. serum reactivity was specific to IBD. The patterns of serum reactivity differed for the different oligopeptides (different sera were reactive with different oligopeptides) suggesting that combining the different candidate biomarkers in a panel may increase diagnostic performance as demonstrated by a preliminary receiver-operator-characteristics (ROC) excluding JOZ240-D7 (Figure 3-11).

Whereas the biomarker panel conferred increased diagnostic performance over single markers with best combined sensitivity and specificity of 84 % and 66.7 %, discrimination between the two phenotypes CD and UC was not possible.

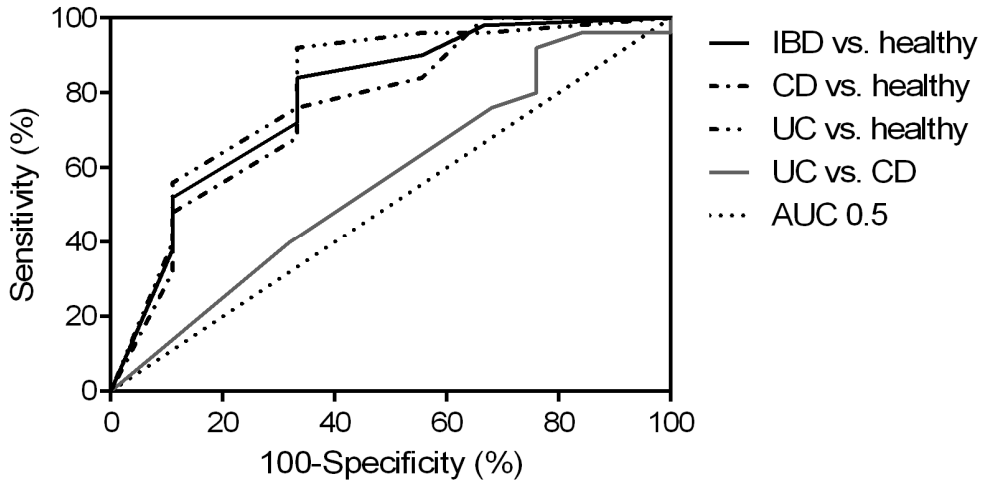


Figure 3-11: Receiver-operator-characteristics for a panel of biomarker candidates (oligopeptide phage). Specificity and sensitivity was calculated at thresholds of 1x, 2x, 3x, 5x, 10x and 20x standard deviation of the negative cohort. If the serum response of at least one marker was greater than the threshold (i.e. positive), the response of the panel was considered positive.

Nevertheless, these results motivated to produce the selected oligopeptides as synthetic peptides or recombinant oligopeptides and full-length proteins to verify these preliminary findings on recombinant protein level and a larger control cohort. Also, full-length proteins would potentially further increase sensitivity due to additional immunogenic epitopes.

3.2.4 Recombinant production of biomarker candidates

For the recombinant production of the selected oligopeptides and the originating full-length proteins (if known), the coding DNA sequences were PCR amplified and cloned into pET21a(+) and pET21a(+)-pelB for both cytoplasmic and periplasmic expression in *E. coli*. Additionally to the oligopeptides selected applying the matched panning approach, clone JOZ191-G8 from the unmatched approach was included in the panel for recombinant protein production. PCR amplification of the genes coding for the full-length proteins was performed using the amplified metagenomic DNA as template and oligonucleotide primers based on the annotations found by blast analysis (Table 3-4 and Table 3-6).

PCR amplification of the genes coding for the full-length protein using metagenomic template DNA resulted in PCR products with the expected fragment size. However, as determined by DNA sequencing, they did not match the found annotation with 100 % identity (Suppl. Table 2 and Suppl. information) suggesting a broad variety of homologue proteins from various species in the gut metaproteome. Cloning the gene coding for the full length protein of JOZ243-F9 (AII76920.1) was not possible at all.

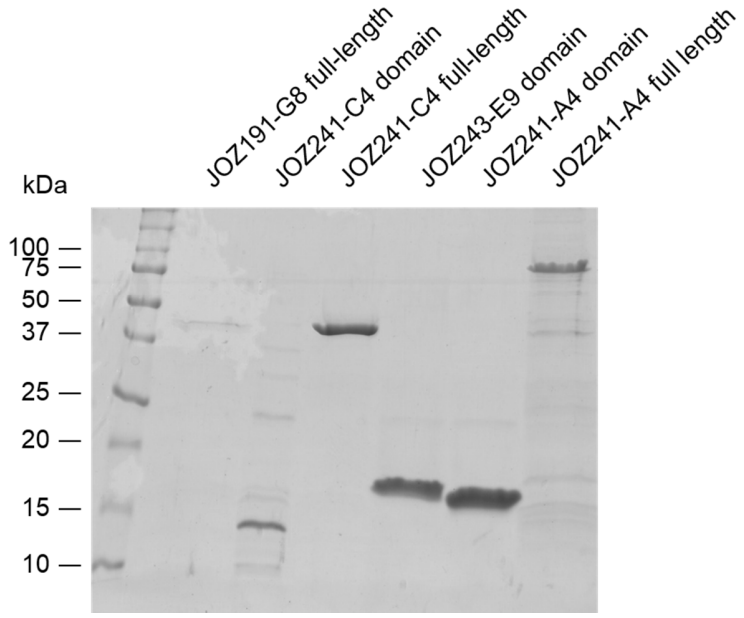


Figure 3-12: Coomassie gel of recombinant biomarker candidate proteins. The full-length protein JOZ191-G8 precipitated after purification and only minute amount of protein were loaded on the gel. The protein preparations of JOZ241-C4 domain and JOZ241-A4 full-length contained marked *E. coli* protein impurities.

The production of the recombinant full-length proteins was only possible for JOZ191-G8, JOZ241-C4 and JOZ241-A4 and for the domains JOZ241-C4, JOZ243-E9 and JOZ241-A4 (Figure 3-12).

The full-length protein JOZ191-G8 precipitated after purification and only minute amounts of proteins remained in solution when 8 M urea was added (Figure 3-12, lane 1, ~40 kDa protein). The protein preparation JOZ241-C4 (domain) and JOZ241-A4 (full-length) contained many *E. coli* protein impurities. The production and purification of the oligopeptides or protein domains of JOZ240-H2, JOZ242-C9, JOZ242-G4 and JOZ243-F9 failed to yield sufficient recombinant protein. Therefore, the oligopeptides JOZ191-G8, JOZ240-H2, JOZ243-F9 and JOZ243-F12 were chemically synthesized as peptides and the clones JOZ242-C9 and JOZ242-G4 had to be excluded from the panel. The resulting panel for further validation as recombinant protein or synthetic peptide is summarized in Table 3-7.

Table 3-7: Summary of biomarker candidates available as recombinant protein or synthetic peptide

Clone	Oligopeptide/ domain	Full-length protein
JOZ191-G8	synthetic	recombinant
JOZ240-H2	synthetic	-
JOZ241-A4	recombinant	recombinant
JOZ241-C4	recombinant	recombinant
JOZ242-C9	failed	failed
JOZ242-G4	failed	-
JOZ243-F12	synthetic	-
JOZ243-E9	recombinant	failed
JOZ243-F9	synthetic	failed

3.2.5 Assessment of diagnostic performance of a biomarker panel (peptide/ recombinant protein)

For the validation of the selected biomarker candidates on recombinant protein level, reactivity of the IBD sera (see above) and healthy control sera was determined by ELISA (Figure 3-13). Additionally to the nine healthy control sera used previously, a larger cohort

of 55 appropriate matched control sera was included (healthy control sera were a kind gift from Dr. Thomas Skurk, TU München, Germany).

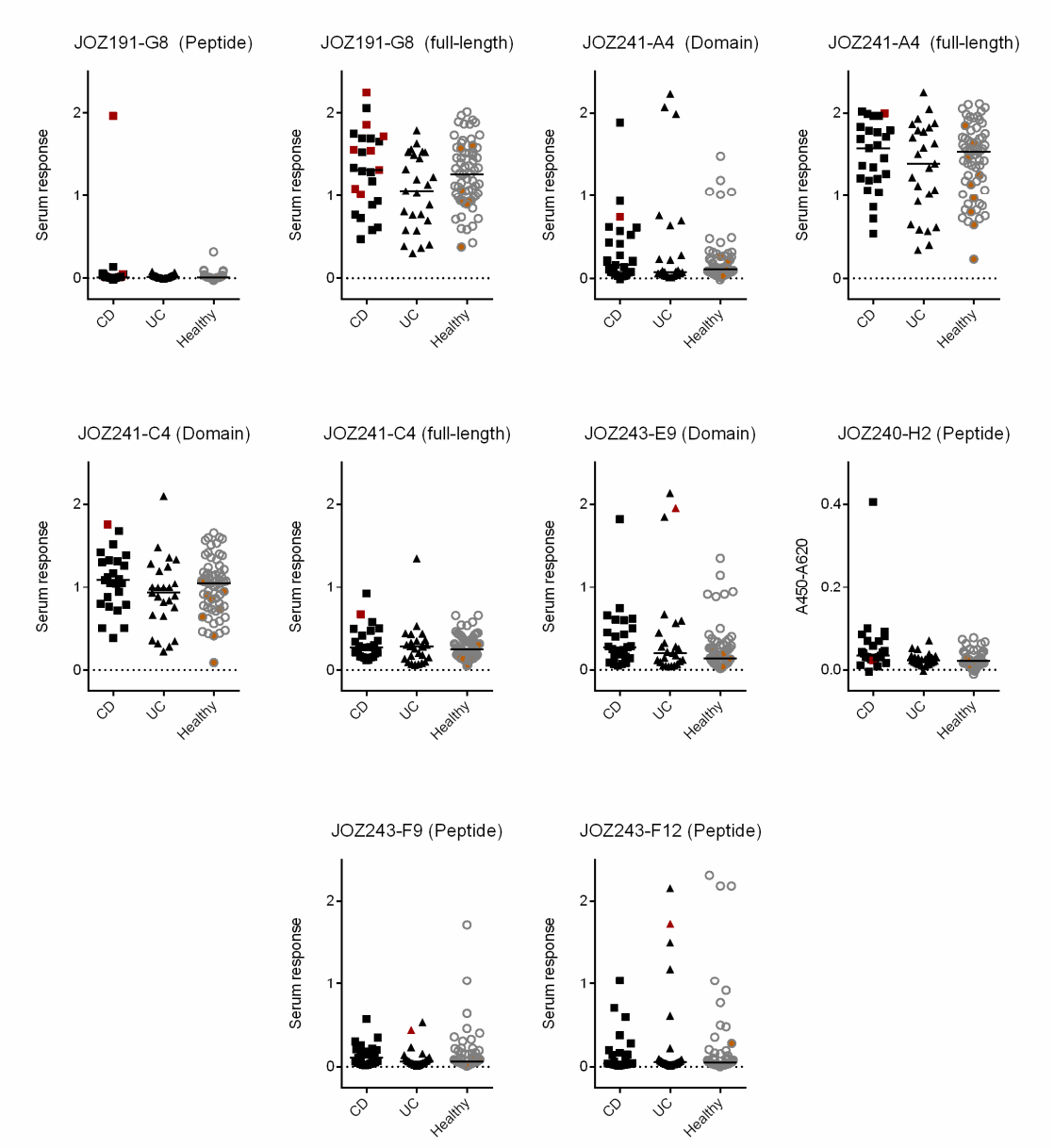


Figure 3-13: Reactivity of IBD serum cohorts with biomarker candidates selected by a matched panning approach (recombinant protein and synthetic peptide). Serum response was determined by ELISA and plotted as mean of triplicated measurement. A background subtraction step was included. The median cohort reactivity is indicated by a black line. Reactivity of the serum (or samples of serum pool) used for candidate selection are highlighted in red; healthy control sera are highlighted in orange. Recombinant proteins and peptides were directly immobilized and bound serum antibodies from single sera (1:500 dilution) were detected with an anti-human (IgA, IgG, IgM) antibody.

The validation of the biomarker candidates on the level of recombinant proteins or synthetic peptides failed. Although some sera of the IBD cohort were reactive with the recombinant proteins and synthetic peptides, an equal portion of sera from the control cohort also reacted with the selected biomarker candidates.

The protein preparations from JOZ191-G8 (full-length), JOZ241-A4 (full-length) and JOZ241-C4 (domain) were most reactive with all sera considered as background and most likely resulted from the high amount of *E. coli* protein impurities in the sample. Background in the remaining samples was decreased especially for the synthetic peptides purified by HPLC.

The serum used for the selection of the biomarker candidates was always (except from JOZ240-H2) among the sera most reactive with the corresponding protein or peptide (Figure 3-13, marked in red) demonstrating the selection was indeed based on the specific interaction of serum antibodies and displayed oligopeptide. Notably, all healthy control sera of the initial validation on oligopeptide phage level were among the least reactive sera (Figure 3-13, marked in orange) recapitulating initial findings that sera from healthy individuals did not react with the biomarker candidates. Nevertheless, these findings were not confirmed with a larger cohort of control sera and consequently the selected biomarker candidates were not specific to IBD.

4 Discussion

The present work describes a biomarker discovery approach based on functional proteomics using the phage display technology (ORFeome phage display). Whereas previous studies successfully identified biomarkers of single pathogens using the ORFeome phage display technology (Becker et al., 2015; Connor et al., 2016; Kügler et al., 2008; Meyer et al., 2012; Naseem et al., 2010), in the present study, the technology was applied for the first time in a metagenomic context. The goal was to identify novel gut microbiome-derived biomarker candidates for inflammatory bowel disease, a chronic inflammatory disorder of the gut that results from an aberrant immune reaction against the gut microbiome in genetically susceptible hosts.

First, as a proof-of-concept, the technology was applied using a mouse model. Therefore, three metaproteome phage display libraries were constructed from gut microbiome-derived metagenomic DNA of $TNF^{\Delta ARE/+}$ mice, a mouse model for Crohn's disease-like ileitis. Successfully identifying three novel immunogenic oligopeptides that had specifically induced an immune response in inflamed mice prove the ORFeome phage display technology to be applicable in a metagenomic context. However, compared to phage display libraries covering genomes of one species, metagenomic approaches require larger libraries. Estimating the necessary total library diversity to cover whole metagenomes is very difficult as the gut microbiome composition and relative abundance of the different species is not completely known. The human gut microbiome harbors several hundred different bacterial species (Human Microbiome Project, 2012) with more than 10 million genes (Li et al., 2014; Qin et al., 2010) and although the libraries contained more than 10^7 clones and up to 20 Gbp, they presumably still did not cover the entire metagenome. Especially genes from species with low abundance may have been underrepresented. Additionally, the metagenomic DNA used for library construction may have also contained DNA of other microorganisms such as fungi and protozoa as well as viruses, host and food-derived DNA. In a previous study, however, serological expression cloning (SEREX) of randomly fragmented metagenomic DNA from the caecum of C3H/HeJBir mice and screening of $6x\ 10^5$ pfu of the lambda expression libraries was sufficient to identify bacterial flagellin as immunogenic protein (Lodes et al., 2004). The metaproteome phage display libraries constructed in the present work contained more

than 10^7 individual clones and exceeded the library size from the study by Lodes et al. at least 20-fold. Moreover, similar to the ORFeome phage display technology, SEREX relies on non-directional cloning of randomly fragmented DNA, so only a small fraction (typically less than 6 %) of inserts usually result in open reading frames. Employing non-lytic M13 phage and a special helperphage with deleted *gIII* ("Hyperphage") (Rondot et al., 2001; Soltes et al., 2007), the ORFeome phage display technology not only allows the screening by monoclonal ELISA but also an ORF enrichment (Hust et al., 2006) prior to selection. The metaproteome phage display libraries of the present work contained up to 80 % open reading frames based on Sanger DNA sequencing of 20 clones. Hence, ORFeome phage display libraries contain a much lower number of junk fragments which can hamper functional selection due to a propagation bias of these fragments (Hust et al., 2006). Although the constructed metaproteome phage libraries did most likely not cover the full gut metagenome, they still had a substantial degree of diversity that allowed their use as starting materials for peptide selection. Moreover, the ability to perform culture-independent proteomic studies is one of the superior characteristics of the technology, clearly demonstrated as in the gut microbiome only 56 and 70 % of 16S rRNA reads detected by sequencing have been related to cultivated bacteria at the species and genus level, respectively (Goodman et al., 2011).

The present study provides the first analysis of ORFeome phage display libraries based on NGS data and revealed that only few clones were overrepresented in the libraries but the majority of sequences were equally distributed. In another NGS study analyzing a commercial peptide phage library (Ph.D.™-12) after amplification, the authors acquired 2x 10^7 reads to analyze a library of 10^6 unique sequences (Matochko et al., 2012). This is about the 50-fold sequencing depth compared to the present study suggesting that the full library diversity of the present study was not covered especially considering that only 10 % of sequences were shared between triplicate library preparations. Nevertheless, sequencing depth was sufficient to gain insights into the size distribution and relative abundance of individual clones in the libraries. Analyzing the Ph.D.™-12 peptide phage library, the authors reported a strong bias towards 150 clones that dominated 20 % of the library (with 20 clones accounting for 8 %) (Matochko et al., 2012). The broad size distribution ranging from 8 to 476 bp would suggest an even stronger bias in the ORFeome

phage display libraries due to different inserts having different impact on cell growth. However, the 20 most abundant clones accounted for less than 0.5 % of total reads, indicating the lack of a respective bias in the ORFeome phage display libraries. Moreover, the size distribution of nucleotide inserts revealed a strong enrichment (70.3 % of total sequences) of clones with inserts of 3n+1 nucleotides representing again a strong enrichment of ORFs as only insert with 3n+1 nucleotides can result in ORFs. The degree of enrichment reflected the ORF enrichment based on small sample DNA sequencing and is in line with previously reported ORF enrichment using the ORFeome phage display technology (Connor et al., 2016; Hust et al., 2006). In the packaged ORFeome phage display libraries remained about 30 % inserts which were not in frame with the pIII and in theory they should not result in infective phage particles. But functionally displayed oligopeptides derived from inserts with frameshifts were described before (Cárcamo et al., 1998; Goldman, Korus, & Mandecki, 2000; Hust et al., 2006) and were proposed to result from RNA secondary structures or a selection pressure against oligopeptides which may be toxic for the *E. coli* host. Therefore, an enrichment for 100 % ORFs may not be possible at all.

For two out of the three selected immunogenic oligopeptides, consisting of 20 and 56 amino acids, an identification by blast analysis was not possible. The third selected oligopeptide comprised 40 amino acids of a phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR synthase) of *Blautia* sp. and is very similar to SAICAR synthases from other highly abundant members of the mammalian gut (e.g. *Dorea* sp. and *Clostridium* sp.) but the exact origin of the DNA sequence was not identified. The closest similarity was to the *Roseburia hominis* SAICAR gene. SAICAR synthase is an enzyme involved in the *de novo* biosynthesis of purine nucleotides (Lukens & Buchanan, 1959) and is essential for all living organisms. Due to structural and functional differences between microbial and vertebrate SAICAR synthase, selective inhibitors have been proposed as antimicrobial agents (Ginder, Binkowski, Fromm, & Honzatko, 2006) and structural differences may confer immunogenicity to bacterial SAICAR synthase.

The antibody response against the selected oligopeptides was not only specific for mice suffering from ileal inflammation but there was also a trend towards a correlation between

ileal inflammation and antibody response. Additionally, mice kept under specific pathogen free conditions that developed less inflammation (and were lacking colonic inflammation) also showed decreased serum response against the selected peptides if present at all. Lacking immune response against the oligopeptide JOZ158-G8 may indicate the absence of the originating species in SPF housing suggesting not only an influence of gut inflammation but also of the gut microbiome composition on the humoral immune response.

The successful identification of immunogenic oligopeptides specific for the $TNF^{\Delta ARE/+}$ mouse model of Crohn's disease using ORFeome phage display motivated to apply the technology also to human IBD samples to identify novel biomarker candidates for IBD. Following two different panning strategies, candidate oligopeptides were selected from gut microbiome-derived metaproteome phage display libraries of IBD patients. However, neither a panning approach that used a pool of IBD sera for selection nor combining library and serum derived from the same donor individual for selection resulted in oligopeptides highly reactive with the majority of samples in the tested IBD cohort. Lacking sensitivity of the selected biomarker candidates suggest that strong individual-specific antibody responses were driving the oligopeptide selection and universal markers may not even be existent with high antibody titers. Nevertheless, the absence of reactive sera in the preliminary control group suggested that antibody responses against gut microbiome-derived antigens are more diverse in humans and a combination of these markers may confer some diagnostic accuracy. Increased diagnostic performance by combining different serologic, genetic and fecal markers has been described before (Plevy et al., 2013). However, the further validation of the selected panel of biomarker candidates was not possible as a larger control group demonstrated the selected oligopeptides were not specific to IBD and even healthy individuals can have high antibody titers against proteins of the gut microbiome.

Despite the transgenic model spontaneously develops gut inflammation and has many shared features with human Crohn's disease such as the involvement of TNF and the gut microbiome as inflammation trigger (Schaubeck et al., 2015), there are certain limitations for the translation to human IBD. The mice used in this study most likely represented a more homogenous cohort regarding the gut microbiome composition than a human IBD

cohort and this might have been reflected in a more homogenous immune response. First of all, host genetics is a determinant of gut microbiome composition. Twin studies in humans revealed the gut microbiome composition in monozygotic twins is more similar than in dizygotic twins (Goodrich et al., 2014). In a study conducted by Turnbaugh and colleagues the trend of less dissimilar microbiomes in monozygotic twins was not significant but the microbiome between twins (di- and monozygotic) was more similar than to their mothers' (Turnbaugh et al., 2009) suggesting that the gut microbiome is shared among family members and the closer host genetics are the more similar the gut microbiota compositions becomes. The transgenic animals used in the present study have a common genetic offspring and partially originated from the same parental animals. Thus, the genetic homogeneity of experimental mice did not reflect the true situation in a human cohort. Further, the maternal transmission of the microbiota in mammals was reported to influence the gut microbiome in mouse studies (Benson et al., 2010; Friswell et al., 2010) and the influence of a common maternal animal has to be considered not only with respect to host genetics. Although some of the mouse cohorts were not housed under strictly controlled SPF conditions, the housing might still have influenced the gut microbiome as cage mates interact with each other. Multiple studies have reported a homogenization of the gut microbiome between co-housed mice (Elinav et al., 2011; Zenewicz et al., 2013) which might result from repeated inoculation of cage mates by coprophagy. Another environmental influence on the gut microbiome is the diet. All mice were fed the same standardized diet and diet was shown to have an impact on gut microbiome composition as it was able to shift the enterotype of wild mice housed under laboratory conditions on chow diet (Wang et al., 2014).

Although all animals in the study had a common genetic background, were fed the same diet and were at the same age there were still inter-individual differences in gut microbiome composition as well as in the reactivity of sera towards the selected oligopeptides. With respect to genetic variability, age and individual lifestyle the gut microbiome may be even more individual-specific in humans. A recent population-level microbiome study has identified 69 factors associated with gut microbiome composition (Falony et al., 2016). Thus, great heterogeneity and relatively small number of human

samples in the cohort may have hampered the selection of universal biomarker candidates.

IBD has been associated with shifts in microbiome composition rather than a single pathogen signature (Gevers et al., 2014). Antibody serum markers with highest prevalence are directed against autoantigens or microbial structures that seem not very specific to single species. ASCAs are antibodies against yeast carbohydrate structures (Sendid et al., 1996) and antibodies against bacterial proteins are directed against flagellin of gram-positive bacteria (Lodes et al., 2004), protein I2 from *Pseudomonas fluorescens*, which is a homologue of bacterial transcription factor tetR (Sutton, Kim, et al., 2000; Wei et al., 2002) as well as *E. coli* OmpC that may be a cross-reaction with autoantigens (Cohavy et al., 2000). Additionally, antibody responses against common gut bacteria were shown to have higher diagnostic accuracy than single antigens (Adams et al., 2008). Consequently, antibodies found in IBD patients potentially result from an increased exposure of gut microbial antigens as a consequence of epithelial leakiness. This might be reflected by the specific antibody response against bacterial SAICAR synthase as a rather universal antigen found in the $TNF^{\Delta ARE/+}$ mouse model. Whether inflammation is causative for barrier breakdown alone still has to be elucidated as findings that affected and unaffected first-degree relatives have increased antibody titers against bacterial antigens suggesting a genetic influence rather than a consequence of inflammation (Mei et al., 2006; Michielan et al., 2013; Sutton, Yang, et al., 2000; Török et al., 2005).

Gut inflammation in $TNF^{\Delta ARE/+}$ mice develops gradually over age (Baur et al., 2011; Kontoyiannis et al., 1999) whereas human IBD is characterized by periodically occurring intestinal inflammation. The samples used for the selection of biomarker candidates were from patients in disease remission. Antibody titers specific for IBD were reported to be stable over the disease (Müller et al., 2005; Seibold, 2005). However, this seems to be controversial as other studies reported a correlation between antibody titers and disease activity (Canani et al., 2004; El-Matary, Dupuis, & Sokoro, 2015) which was supported by the findings that serum response in $TNF^{\Delta ARE/+}$ mice tended to correlate with gut inflammation. Moreover, the successful identification of bacterial flagellin as serologic marker for Crohn's disease (Lodes et al., 2004) was based on a colitogenic mouse model that develops strong IgG responses against antigens of the enteric flora (Brandwein et al.,

1997). The ASCA/ANCA status as surrogate for antibody titers against microbiota-derived antigens was unknown for the human IBD samples and information about anti-TNF treatment was not available. Anti-TNF treatment was previously reported to impact ASCA levels (Frehn et al., 2014). Therefore, from a technological perspective the relevant antibody titers might not have been high enough to allow the selection of disease-specific biomarkers. Heterogeneity of the human IBD cohorts together with phenotypic differences in disease course compared to the mouse model added another level of complexity to the selection of biomarker candidates. Consequently, as long as the complex pathogenic mechanisms underlying IBD are not completely understood, it might be difficult to develop a rational strategy for the selection of novel biomarker candidates.

In conclusion, despite the ORFeome phage display technology has demonstrated promising features for the identification of microbiome-derived biomarkers in a mouse model of Crohn's disease it was not possible to eventually identify such markers using human IBD samples. Heterogeneity of the human IBD cohorts as well as unclear antibody levels in the IBD samples may have hampered biomarker candidate selection. Nevertheless, the technology was proven for its applicability to metagenomic samples and libraries may be constructed from any type of genomic or metagenomic DNA. Thus, ORFeome phage display may also facilitate the use of environmental metagenomes to identify protein interaction partners of microbial communities with potential further interest for biotechnological or industrial processes.

5 Outlook

The present study described the first successful application of ORFeome phage display in a metagenomic context and prove the technology to be applicable to study host-microbiome interactions. Metagenomes provide a rich source for novel protein functions and the technology may not only be used to identify disease related biomarkers but also protein interaction partners from environmental samples that are of biotechnological interest.

Further technological development of ORFeome phage display may still be possible considering library design. Using a more homogenous insert size distribution in initial libraries may decrease bias towards smaller fragments during ORF enrichment and phage packaging. With respect to the increased complexity when using whole metagenomes for library construction, the ELISA-based screening approach of 92 clones may limit the panning output especially considering the heterogeneity of human cohorts in a disease context like IBD. Future studies may include NGS analysis to provide a more comprehensive analysis of panning output and array-based screening methods may identify profiles of antibody responses rather than single immunogenic proteins. Using newly developed array copier and *in vitro* translation technologies, array-based screening might be implemented into the workflow and make the ORFeome phage display technology even more powerful.

6 Summary

Pathogen infections, autoimmune diseases, and chronic inflammatory disorders are associated with systemic antibody responses from the host immune system. Disease-specific antibodies can be important serum biomarkers, but the identification of antigens associated with specific immune reactions is challenging, in particular if complex communities of microorganisms are responsible for the disease. Despite promising new diagnostic opportunities, the discovery of these serological markers is difficult and becomes even more challenging with increasing complexity of the microbial communities.

In the present work, a metagenomic M13 phage display approach (ORFeome phage display) was used to identify novel gut microbiome-derived biomarker candidates in the disease context of inflammatory bowel disease (IBD). As a proof-of-concept, $TNF^{\Delta ARE/+}$ mice of Crohn's-like experimental ileitis were used. Three individual metaproteome phage display libraries with a library size of approximately 10^7 clones each were constructed from the caecum of ileal inflamed mice. The libraries were analyzed by next generation sequencing revealing a substantial degree of diversity for oligopeptide selection. Using serum antibodies, three oligopeptides that induced specific antibody responses in inflamed mice were selected and validated.

Translating the technology to human IBD, metaproteome phage display libraries were constructed from the gut microbiome of patients suffering from Crohn's disease and ulcerative colitis. Several candidate oligopeptides were selected using patient sera and were initially validated on oligopeptide phage level with a small control cohort. But using a larger control cohort, it was not possible to confirm the selected immunogenic proteins being IBD-specific. Broader cohort heterogeneity of humans compared to a mouse model may have hampered the selection of specific biomarker candidates for human IBD. Nevertheless, the present work provides the first successful application of ORFeome phage display using a metaproteome approach for the study of protein-protein interactions and the discovery of potential disease biomarkers.

7 Acknowledgements

The last years were a journey full of excitement and frustration. But every single experience was valuable and let me broaden my horizon in many ways. This fantastic trip was accompanied by several persons who contributed in one or another way and I want to express my appreciation.

First of all, my deepest gratitude goes to my mentor and first examiner of this thesis, **Prof. Dr. Michael Hust**, for his scientific guidance and who has always been supportive throughout the last years. He found the perfect balance between the focus on this study and the freedom that allowed me to follow other interesting topics which were not related to this study but a huge benefit to my professional career.

I am also very thankful to **Prof. Dr. Stefan Dübel**, for being the second examiner of this thesis, for the scientific feedback and for facilitating the participation at numerous conferences as well as the support of all my “extra-curricular” activities which were crucial to make this journey an overall success story.

My gratitude also goes to **Prof. Dr. Susanne Engelmann** for being third examiner of this thesis and **Prof. Dr. André Fleißner** who was so kind to be head of the thesis committee.

I want to thank **Dr. habil. Thomas Clavel**, my major collaborator, for his contribution with material, network and knowledge. I very much appreciated his input, the scientific discussions and critical feedback. Also thanks to **Dr. Ilias Lagkouvardos** for the excellent NGS work and everyone in Thomas’ team, especially **Sarah Just**, who gave me a warm welcome during a guest stay at TUM and supported me with all necessary information about the mouse work.

This study was realized in a collaborative project and would not have been possible without the crucial contribution of **Dr. Patricia Lepage**, **Dr. Harry Sokol**, **Prof. Dr. Hans Hauner** and **Dr. Thomas Skurk** who were so kind to provide precious material. Thank you very much!

Being a PhD student often means spending long hours in the lab with bad data at the end of the day. Therefore, I want to thank **all members of the Department of Biotechnology** for the great work environment which helped to overcome frustrating episodes. Thanks to

my colleagues for the occasional after-work beer, probably the most productive brain-storming time (but optional). I want to thank **Saskia Helmsing** for excellent technical lab support, **Philipp Kuhn** for the daily discussions in the office and **Tobias Unkauf** for everything we achieved as a team during the last years. Thanks to my master students, **Felix Pritsch**, **Justyna Gerth** and **Hanna Lunding** and the whole **iGEM Team Braunschweig 2013** for an incredible time.

Last but not least, I want to thank my family and especially my parents, **Monika Fink-Zantow** and **Christoph Zantow**, for continuous support throughout my studies and who enabled to me to go this way. This thesis is dedicated to my grandma **Maria Fink**, who used to call me “doctor” ever since.

8 List of Figures

9 List of Tables

10 References

Adams, R. J., Heazlewood, S. P., Gilshenan, K. S., O'Brien, M., McGuckin, M. A., & Florin, T. H. (2008). IgG antibodies against common gut bacteria are more diagnostic for Crohn's disease than IgG against mannan or flagellin. *Am J Gastroenterol, 103*(2), 386-396.

Ballew, J. T., Murray, J. A., Collin, P., Mäki, M., Kagnoff, M. F., Kaukinen, K., & Daugherty, P. S. (2013). Antibody biomarker discovery through in vitro directed evolution of consensus recognition epitopes. *Proc Natl Acad Sci U S A, 110*(48), 19330-19335.

Baur, P., Martin, F. P., Gruber, L., Bosco, N., Brahmbhatt, V., Collino, S., Guy, P., Montoliu, I., Rozman, J., Klingenspor, M., Tavazzi, I., Thorimbert, A., Rezzi, S., Kochhar, S., Benyacoub, J., Kollias, G., & Haller, D. (2011). Metabolic phenotyping of the Crohn's disease-like IBD etiopathology in the TNF(DeltaARE/WT) mouse model. *J Proteome Res, 10*(12), 5523-5535.

Becker, M., Felsberger, A., Frenzel, A., Shattuck, W. M., Dyer, M., Kügler, J., Zantow, J., Mather, T. N., & Hust, M. (2015). Application of M13 phage display for identifying immunogenic proteins from tick (Ixodes scapularis) saliva. *BMC Biotechnol, 15*(1).

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., Zhang, M., Oh, P. L., Nehrenberg, D., Hua, K., Kachman, S. D., Moriyama, E. N., Walter, J., Peterson, D. A., & Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences, 107*(44), 18933-18938.

Beranova-Giorgianni, S. (2003). Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. *TrAC Trends in Analytical Chemistry, 22*(5), 273-281.

Bernstein, C. N. (2015). Treatment of IBD: Where We Are and Where We Are Going. *Am J Gastroenterol, 110*(1), 114-126.

Blüthner, M., Bautz, E. K., & Bautz, F. A. (1996). Mapping of epitopes recognized by PM/Scl autoantibodies with gene-fragment phage display libraries. *J Immunol Methods, 198*(2), 187-198.

Bonneau, J., Dumestre-Perard, C., Rinaudo-Gaujous, M., Genin, C., Sparrow, M., Roblin, X., & Paul, S. (2015). Systematic review: New serological markers (anti-glycan, anti-GP2, anti-GM-CSF Ab) in the prediction of IBD patient outcomes. *Autoimmunity Reviews, 14*(3), 231-245.

Bradbury, A. R., Sidhu, S., Dübel, S., & McCafferty, J. (2011). Beyond natural antibodies: the power of in vitro display technologies. *Nat Biotechnol, 29*(3), 245-254.

Brandwein, S. L., McCabe, R. P., Cong, Y., Waites, K. B., Ridwan, B. U., Dean, P. A., Ohkusa, T., Birkenmeier, E. H., Sundberg, J. P., & Elson, C. O. (1997). Spontaneously colitic C3H/HeJBir mice demonstrate selective antibody reactivity to antigens of the enteric bacterial flora. *J Immunol, 159*(1), 44-52.

Breitling, F., Dübel, S., Seehaus, T., Klewinghaus, I., & Little, M. (1991). A surface expression vector for antibody screening. *Gene, 104*(2), 147-153.

Burisch, J., Jess, T., Martinato, M., & Lakatos, P. L. (2013). The burden of inflammatory bowel disease in Europe. *J Crohns Colitis, 7*(4), 322-337.

Canani, R. B., Romano, M. T., Greco, L., Terrin, G., Sferlazzas, C., Barabino, A., Fontana, M., Roggero, P., Guariso, G., De Angelis, G., Fecarotta, S., Polito, G., & Cucchiara, S. (2004). Effects of disease activity on anti-Saccharomyces cerevisiae antibodies:

implications for diagnosis and follow-up of children with Crohn's disease. *Inflamm Bowel Dis, 10*(3), 234-239.

Cárcamo, J., Ravera, M. W., Brissette, R., Dedova, O., Beasley, J. R., Alam-Moghé, A., Wan, C., Blume, A., & Mandecki, W. (1998). Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library. *Proc Natl Acad Sci U S A, 95*(19), 11146-11151.

Chen, C. S., Sullivan, S., Anderson, T., Tan, A. C., Alex, P. J., Brant, S. R., Cuffari, C., Bayless, T. M., Talor, M. V., Burek, C. L., Wang, H., Li, R., Datta, L. W., Wu, Y., Winslow, R. L., Zhu, H., & Li, X. (2009). Identification of novel serological biomarkers for inflammatory bowel disease using Escherichia coli proteome chip. *Mol Cell Proteomics, 8*(8), 1765-1776.

Chen, Y. T., Gure, A. O., & Scanlan, M. J. (2005). Serological analysis of expression cDNA libraries (SEREX): an immunoscreening technique for identifying immunogenic tumor antigens. *Methods Mol Med, 103*, 207-216.

Ciric, M., Moon, C. D., Leahy, S. C., Creevey, C. J., Altermann, E., Attwood, G. T., Rakonjac, J., & Gagic, D. (2014). Metasecretome-selective phage display approach for mining the functional potential of a rumen microbial community. *BMC Genomics, 15*, 356.

Cohavy, O., Bruckner, D., Gordon, L. K., Misra, R., Wei, B., Eggena, M. E., Targan, S. R., & Braun, J. (2000). Colonic bacteria express an ulcerative colitis pANCA-related protein epitope. *Infection and Immunity, 68*(3), 1542-1548.

Colman, R. J., & Rubin, D. T. (2014). Fecal microbiota transplantation as therapy for inflammatory bowel disease: A systematic review and meta-analysis. *Journal of Crohn's and Colitis, 8*(12), 1569-1581.

Connor, D. O., Zantow, J., Hust, M., Bier, F. F., & von Nickisch-Rosenegk, M. (2016). Identification of Novel Immunogenic Proteins of Neisseria gonorrhoeae by Phage Display. *PLoS ONE, 11*(2), e0148986.

Cosnes, J., Gower-Rousseau, C., Seksik, P., & Cortot, A. (2011). Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology, 140*(6), 1785-1794.

Crameri, R., & Walter, G. (1999). Selective enrichment and high-throughput screening of phage surface-displayed cDNA libraries from complex allergenic systems. *Comb Chem High Throughput Screen, 2*(2), 63-72.

Danckert, L., Hoppe, S., Bier, F. F., & von Nickisch-Rosenegk, M. (2014). Rapid identification of novel antigens of Enteritidis by microarray-based immunoscreening. *Mikrochim Acta, 181*(13-14), 1707-1714.

Danese, S., Fiorino, G., Peyrin-Biroulet, L., Lucenteforte, E., Virgili, G., Moja, L., & Bonovas, S. (2014). Biological agents for moderately to severely active ulcerative colitis: a systematic review and network meta-analysis. *Ann Intern Med, 160*(10), 704-711.

Delvecchio, V. G., Connolly, J. P., Alefantis, T. G., Walz, A., Quan, M. A., Patra, G., Ashton, J. M., Whittington, J. T., Chafin, R. D., Liang, X., Grewal, P., Khan, A. S., & Mujer, C. V. (2006). Proteomic profiling and identification of immunodominant spore antigens of Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis. *Appl Environ Microbiol, 72*(9), 6355-6363.

Dotan, I., Fishman, S., Dgani, Y., Schwartz, M., Karban, A., Lerner, A., Weishauss, O., Spector, L., Shtevi, A., Altstock, R. T., Dotan, N., & Halpern, Z. (2006). Antibodies

Against Laminaribioside and Chitobioside Are Novel Serologic Markers in Crohn's Disease. *Gastroenterology, 131*(2), 366-378.

Dotan, N., Altstock, R. T., Schwarz, M., & Dukler, A. (2006). Anti-glycan antibodies as biomarkers for diagnosis and prognosis. *Lupus, 15*(7), 442-450.

Dübel, S., Stoevesandt, O., Taussig, M. J., & Hust, M. (2010). Generating recombinant antibodies to the complete human proteome. *Trends Biotechnol, 28*(7), 333-339.

Dubinsky, M. C., Kugathasan, S., Mei, L., Picornell, Y., Nebel, J., Wrobel, I., Quiros, A., Silber, G., Wahbeh, G., Katzir, L., Vasiliauskas, E., Bahar, R., Otley, A., Mack, D., Evans, J., Rosh, J., Hemker, M. O., Leleiko, N., Crandall, W., Langton, C., Landers, C., Taylor, K. D., Targan, S. R., Rotter, J. I., Markowitz, J., & Hyams, J. (2008). Increased immune reactivity predicts aggressive complicating Crohn's disease in children. *Clin Gastroenterol Hepatol, 6*(10), 1105-1111.

Duck, L. W., Walter, M. R., Novak, J., Kelly, D., Tomasi, M., Cong, Y., & Elson, C. O. (2007). Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis, 13*(10), 1191-1201.

Duerr, R. H., Taylor, K. D., Brant, S. R., Rioux, J. D., Silverberg, M. S., Daly, M. J., Steinhart, A. H., Abraham, C., Regueiro, M., Griffiths, A., Dassopoulos, T., Bitton, A., Yang, H., Targan, S., Datta, L. W., Kistner, E. O., Schumm, L. P., Lee, A. T., Gregersen, P. K., Barmada, M. M., Rotter, J. I., Nicolae, D. L., & Cho, J. H. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science, 314*(5804), 1461-1463.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.

El-Matary, W., Dupuis, K., & Sokoro, A. (2015). Anti-Saccharomyces cerevisiae antibody titres correlate well with disease activity in children with Crohn's disease. *Acta Paediatr, 104*(8), 827-830.

Elinav, E., Strowig, T., Kau, A. L., Henao-Mejia, J., Thaiss, C. A., Booth, C. J., Peaper, D. R., Bertin, J., Eisenbarth, S. C., Gordon, J. I., & Flavell, R. A. (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell, 145*(5), 745-757.

Faix, P. H., Burg, M. A., Gonzales, M., Ravey, E. P., Baird, A., & Larocca, D. (2004). Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. *Biotechniques, 36*(6), 1018-1022, 1024, 1026-1019.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M. J., Valles-Colomer, M., Vandeputte, D., Tito, R. Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G., D'Hoe, K., Jonckheere, K., Homola, D., Garcia, R., Tigchelaar, E. F., Eeckhaudt, L., Fu, J., Henckaerts, L., Zhernakova, A., Wijmenga, C., & Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science, 352*(6285), 560-564.

Fehrsen, J., & du Plessis, D. H. (1999). Cross-reactive epitope mimics in a fragmented-genome phage display library derived from the rickettsia, Cowdria ruminantium. *Immunotechnology, 4*(3-4), 175-184.

Ferrante, M., Henckaerts, L., Joossens, M., Pierik, M., Joossens, S., Dotan, N., Norman, G. L., Altstock, R. T., Van Steen, K., Rutgeerts, P., Van Assche, G., & Vermeire, S. (2007). New serological markers in inflammatory bowel disease are associated with complicated disease behaviour. *Gut, 56*(10), 1394-1403.

Frehn, L., Jansen, A., Bennek, E., Mandic, A. D., Temizel, I., Tischendorf, S., Verdier, J., Tacke, F., Streetz, K., Trautwein, C., & Sellge, G. (2014). Distinct patterns of IgG and IgA against food and microbial antigens in serum and feces of patients with inflammatory bowel diseases. *PLoS ONE, 9*(9), e106750.

Friswell, M. K., Gika, H., Stratford, I. J., Theodoridis, G., Telfer, B., Wilson, I. D., & McBain, A. J. (2010). Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. *PLoS ONE, 5*(1), e8584.

Fuchs, M., Kämpfer, S., Helmsing, S., Spallek, R., Oehlmann, W., Prilop, W., Frank, R., Dübel, S., Singh, M., & Hust, M. (2014). Novel human recombinant antibodies against Mycobacterium tuberculosis antigen 85B. *BMC Biotechnol, 14*, 68.

Gevers, D., Kugathasan, S., Denson, Lee A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, Se J., Yassour, M., Morgan, Xochitl C., Kostic, Aleksandar D., Luo, C., González, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., Stephens, M., Heyman, M., Markowitz, J., Baldassano, R., Griffiths, A., Sylvester, F., Mack, D., Kim, S., Crandall, W., Hyams, J., Huttenhower, C., Knight, R., & Xavier, Ramnik J. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe, 15*(3), 382-392.

Ginder, N. D., Binkowski, D. J., Fromm, H. J., & Honzatko, R. B. (2006). Nucleotide complexes of Escherichia coli phosphoribosylaminoimidazole succinocarboxamide synthetase. *J Biol Chem, 281*(30), 20680-20688.

Goldman, E., Korus, M., & Mandecki, W. (2000). Efficiencies of translation in three reading frames of unusual non-ORF sequences isolated from phage display. *FASEB J, 14*(3), 603-611.

Gonzalez, E., Robles, Y., Govezensky, T., Bobes, R. J., Gevorkian, G., & Manoutcharian, K. (2010). Isolation of neurocysticercosis-related antigens from a genomic phage display library of Taenia solium. *J Biomol Screen, 15*(10), 1268-1273.

Goodman, A. L., Kallstrom, G., Faith, J. J., Reyes, A., Moore, A., Dantas, G., & Gordon, J. I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A, 108*(15), 6252-6257.

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. T., Spector, T. D., Clark, A. G., & Ley, R. E. (2014). Human genetics shape the gut microbiome. *Cell, 159*(4), 789-799.

Goyal, N., Rana, A., Ahlawat, A., Bijjem, K. R., & Kumar, P. (2014). Animal models of inflammatory bowel disease: a review. *Inflammopharmacology, 22*(4), 219-233.

Han, X., Uchida, K., Jurickova, I., Koch, D., Willson, T., Samson, C., Bonkowski, E., Trauernicht, A., Kim, M. O., Tomer, G., Dubinsky, M., Plevy, S., Kugathsan, S., Trapnell, B. C., & Denson, L. A. (2009). Granulocyte-macrophage colony-stimulating factor autoantibodies in murine ileitis and progressive ileal Crohn's disease. *Gastroenterology, 136*(4), 1261-1271, e1261-1263.

Helb, D. A., Tetteh, K. K., Felgner, P. L., Skinner, J., Hubbard, A., Arinaitwe, E., Mayanja-Kizza, H., Ssewanyana, I., Kamya, M. R., Beeson, J. G., Tappero, J., Smith, D. L., Crompton, P. D., Rosenthal, P. J., Dorsey, G., Drakeley, C. J., & Greenhouse, B. (2015). Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proc Natl Acad Sci U S A, 112*(32), E4438-4447.

Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol, 23*(9), 1105-1116.

Hoppe, S., Bier, F. F., & von Nickisch-Rosenegk, M. (2013). Rapid identification of novel immunodominant proteins and characterization of a specific linear epitope of Campylobacter jejuni. *PLoS ONE, 8*(5), e65837.

Hörmannsperger, G., Schaubeck, M., & Haller, D. (2015). Intestinal Microbiota in Animal Models of Inflammatory Diseases. *Ilar j, 56*(2), 179-191.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sahbatou, M., & Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature, 411*(6837), 599-603.

Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature, 486*(7402), 207-214.

Hust, M., Meysing, M., Schirrmann, T., Selke, M., Meens, J., Gerlach, G. F., & Dübel, S. (2006). Enrichment of open reading frames presented on bacteriophage M13 using hyperphage. *Biotechniques, 41*(3), 335-342.

Iskandar, H. N., & Ciorba, M. A. (2012). Biomarkers in inflammatory bowel disease: current practices and recent advances. *Transl Res, 159*(4), 313-325.

Jacobsen, I. D., Meens, J., Baltes, N., & Gerlach, G. F. (2005). Differential expression of non-cytoplasmic Actinobacillus pleuropneumoniae proteins induced by addition of bronchoalveolar lavage fluid. *Vet Microbiol, 109*(3-4), 245-256.

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, L. P., Sharma, Y., Anderson, C. A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S. L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J. P., Ahmad, T., Amininejad, L., Ananthakrishnan, A. N., Andersen, V., Andrews, J. M., Baidoo, L., Balschun, T., Bampton, P. A., Bitton, A., Boucher, G., Brand, S., Buning, C., Cohain, A., Cichon, S., D'Amato, M., De Jong, D., Devaney, K. L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L. R., Franchimont, D., Fransen, K., Gearry, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X., Karlsen, T. H., Kupcinskas, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I. C., Lees, C. W., Louis, E., Mahy, G., Mansfield, J., Morgan, A. R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C. Y., Potocnik, U., Prescott, N. J., Regueiro, M., Rotter, J. I., Russell, R. K., Sanderson, J. D., Sans, M., Satsangi, J., Schreiber, S., Simms, L. A., Sventoraityte, J., Targan, S. R., Taylor, K. D., Tremelling, M., Verspaget, H. W., De Vos, M., Wijmenga, C., Wilson, D. C., Winkelmann, J., Xavier, R. J., Zeissig, S., Zhang, B., Zhang, C. K., Zhao, H., Silverberg, M. S., Annese, V., Hakonarson, H., Brant, S. R., Radford-Smith, G., Mathew, C. G., Rioux, J. D., Schadt, E. E., Daly, M. J., Franke, A., Parkes, M., Vermeire, S., Barrett, J. C., & Cho, J. H. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature, 491*(7422), 119-124.

Kappelman, M. D., Moore, K. R., Allen, J. K., & Cook, S. F. (2013). Recent trends in the prevalence of Crohn's disease and ulcerative colitis in a commercially insured US population. *Dig Dis Sci, 58*(2), 519-525.

Katakura, K., Lee, J., Rachmilewitz, D., Li, G., Eckmann, L., & Raz, E. (2005). Toll-like receptor 9-induced type I IFN protects mice from experimental colitis. *J Clin Invest, 115*(3), 695-702.

Kevans, D., Waterman, M., Milgrom, R., Xu, W., Van Assche, G., & Silverberg, M. (2015). Serological markers associated with disease behavior and response to anti-tumor necrosis factor therapy in ulcerative colitis. *J Gastroenterol Hepatol, 30*(1), 64-70.

Khan, I. H., Ravindran, R., Krishnan, V. V., Awan, I. N., Rizvi, S. K., Saqib, M. A., Shahzad, M. I., Tahseen, S., Ireton, G., Goulding, C. W., Felgner, P., DeRiemer, K., Khanum, A., & Luciw, P. A. (2011). Plasma antibody profiles as diagnostic biomarkers for tuberculosis. *Clin Vaccine Immunol, 18*(12), 2148-2153.

Khan, K. J., Ullman, T. A., Ford, A. C., Abreu, M. T., Abadir, A., Marshall, J. K., Talley, N. J., & Moayyedi, P. (2011). Antibiotic therapy in inflammatory bowel disease: a systematic review and meta-analysis. *Am J Gastroenterol, 106*(4), 661-673.

Kodzius, R., Rhyner, C., Konthur, Z., Buczek, D., Lehrach, H., Walter, G., & Crameri, R. (2003). Rapid identification of allergen-encoding cDNA clones by phage display and high-density arrays. *Comb Chem High Throughput Screen, 6*(2), 147-154.

Kontoyiannis, D., Pasparakis, M., Pizarro, T. T., Cominelli, F., & Kollias, G. (1999). Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity, 10*(3), 387-398.

Kügler, J., Nieswandt, S., Gerlach, G. F., Meens, J., Schirrmann, T., & Hust, M. (2008). Identification of immunogenic polypeptides from a Mycoplasma hyopneumoniae genome library by phage display. *Applied Microbiology and Biotechnology, 80*(3), 447-458.

Kunnath-Velayudhan, S., Salamon, H., Wang, H. Y., Davidow, A. L., Molina, D. M., Huynh, V. T., Cirillo, D. M., Michel, G., Talbot, E. A., Perkins, M. D., Felgner, P. L., Liang, X., & Gennaro, M. L. (2010). Dynamic antibody responses to the Mycobacterium tuberculosis proteome. *Proc Natl Acad Sci U S A, 107*(33), 14703-14708.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680-685.

LaFrentz, B. R., LaPatra, S. E., Call, D. R., Wiens, G. D., & Cain, K. D. (2011). Identification of immunogenic proteins within distinct molecular mass fractions of Flavobacterium psychrophilum. *J Fish Dis, 34*(11), 823-830.

Lagkouvardos, I., Kläring, K., Heinzmann, S. S., Platz, S., Scholz, B., Engel, K. H., Schmitt-Kopplin, P., Haller, D., Rohn, S., Skurk, T., & Clavel, T. (2015). Gut metabolites and bacterial community networks during a pilot intervention study with flaxseeds in healthy adult men. *Mol Nutr Food Res, 59*(8), 1614-1628.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J. R., Prifti, E., Nielsen, T., Juncker, A. S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J. Y., Edris, S., Yang, H., Wang, J., Hansen, T., Nielsen, H. B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Dore, J., Ehrlich, S. D., Bork, P., & Wang, J. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol, 32*(8), 834-841.

Liu, S., Han, W., Sun, C., Lei, L., Feng, X., Yan, S., Diao, Y., Gao, Y., Zhao, H., Liu, Q., Yao, C., & Li, M. (2011). Subtractive screening with the Mycobacterium

tuberculosis surface protein phage display library. *Tuberculosis (Edinb), 91*(6), 579-586.

Lodes, M. J., Cong, Y., Elson, C. O., Mohamath, R., Landers, C. J., Targan, S. R., Fort, M., & Hershberg, R. M. (2004). Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest, 113*(9), 1296-1306.

Lukens, L. N., & Buchanan, J. M. (1959). Biosynthesis of the purines. XXIII. The enzymatic synthesis of N-(5-amino-1-ribosyl-4-imidazolylcarbonyl)-L-aspartic acid 5'-phosphate. *J Biol Chem, 234*(7), 1791-1798.

Main, J., McKenzie, H., Yeaman, G. R., Kerr, M. A., Robson, D., Pennington, C. R., & Parratt, D. (1988). Antibody to Saccharomyces cerevisiae (bakers' yeast) in Crohn's disease. *British Medical Journal, 297*(6656), 1105-1106.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., & Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut, 55*(2), 205-211.

Matochko, W. L., Chu, K., Jin, B., Lee, S. W., Whitesides, G. M., & Derda, R. (2012). Deep sequencing analysis of phage libraries using Illumina platform. *Methods, 58*(1), 47-55.

Mazumdar, S. (2009). Raxibacumab. *MAbs, 1*(6), 531-538.

McCafferty, J., Griffiths, A. D., Winter, G., & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature, 348*(6301), 552-554.

Mehta, F. (2016). Report: economic implications of inflammatory bowel disease and its management. *Am J Manag Care, 22*(3 Suppl), s51-60.

Mei, L., Targan, S. R., Landers, C. J., Dutridge, D., Ippoliti, A., Vasiliauskas, E. A., Papadakis, K. A., Fleshner, P. R., Rotter, J. I., & Yang, H. (2006). Familial expression of anti-Escherichia coli outer membrane porin C in relatives of patients with Crohn's disease. *Gastroenterology, 130*(4), 1078-1085.

Meyer, T., Schirrmann, T., Frenzel, A., Miethe, S., Stratmann-Selke, J., Gerlach, G. F., Strutzberg-Minder, K., Dübel, S., & Hust, M. (2012). Identification of immunogenic proteins and generation of antibodies against Salmonella Typhimurium using phage display. *BMC Biotechnol, 12*, 29.

Michielan, A., Basso, D., Martinato, M., Pathak, S., Banerjee, A., Oliva, L., Plebani, M., Sturniolo, G. C., & D'Incà, R. (2013). Increased antibody response to microbial antigens in patients with Crohn's disease and their unaffected first-degree relatives. *Digestive and Liver Disease, 45*(11), 894-898.

Miltiadou, D. R., Mather, A., Vilei, E. M., & Du Plessis, D. H. (2009). Identification of genes coding for B cell antigens of Mycoplasma mycoides subsp. mycoides Small Colony (MmmSC) by using phage display. *BMC Microbiol, 9*, 215.

Mitsuyama, K., Niwa, M., Masuda, J., Kuwaki, K., Yamasaki, H., Takedatsu, H., Kobayashi, T., & Sata, M. (2011). Isolation and characterization of a novel short peptide associated with Crohn's disease. *Clin Exp Immunol, 166*(1), 72-79.

Moayyedi, P. (2016). Fecal transplantation: any real hope for inflammatory bowel disease? *Curr Opin Gastroenterol, 32*(4), 282-286.

Moayyedi, P., Surette, M. G., Kim, P. T., Libertucci, J., Wolfe, M., Onischi, C., Armstrong, D., Marshall, J. K., Kassam, Z., Reinisch, W., & Lee, C. H. (2015). Fecal Microbiota

Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. *Gastroenterology, 149*(1), 102-109.e106.

Molodecky, N. A., Soon, I. S., Rabi, D. M., Ghali, W. A., Ferris, M., Chernoff, G., Benchimol, E. I., Panaccione, R., Ghosh, S., Barkema, H. W., & Kaplan, G. G. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology, 142*(1), 46-54.e42; quiz e30.

Mow, W. S., Vasiliauskas, E. A., Lin, Y. C., Fleshner, P. R., Papadakis, K. A., Taylor, K. D., Landers, C. J., Abreu-Martin, M. T., Rotter, J. I., Yang, H., & Targan, S. R. (2004). Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology, 126*(2), 414-424.

Müller, S., Styner, M., Seibold-Schmid, B., Flogerzi, B., Mähler, M., Konrad, A., & Seibold, F. (2005). Anti-Saccharomyces cerevisiae antibody titers are stable over time in Crohn's patients and are not inducible in murine models of colitis. *World J Gastroenterol, 11*(44), 6988-6994.

Naseem, S., Meens, J., Jores, J., Heller, M., Dübel, S., Hust, M., & Gerlach, G. F. (2010). Phage display-based identification and potential diagnostic application of novel antigens from Mycoplasma mycoides subsp. mycoides small colony type. *Veterinary Microbiology, 142*(3-4), 285-292.

Nguyen, D. L., Nguyen, E. T., & Bechtold, M. L. (2015). pANCA positivity predicts lower clinical response to infliximab therapy among patients with IBD. *South Med J, 108*(3), 139-143.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods, 8*(10), 785-786.

Peyrin-Biroulet, L. (2010). Anti-TNF therapy in inflammatory bowel diseases: a huge review. *Minerva Gastroenterol Dietol, 56*(2), 233-243.

Peyrin-Biroulet, L., Deltenre, P., de Suray, N., Branche, J., Sandborn, W. J., & Colombel, J. F. (2008). Efficacy and safety of tumor necrosis factor antagonists in Crohn's disease: meta-analysis of placebo-controlled trials. *Clin Gastroenterol Hepatol, 6*(6), 644-653.

Peyrin-Biroulet, L., Standaert-Vitse, A., Branche, J., & Chamaillard, M. (2007). IBD serological panels: Facts and perspectives. *Inflamm Bowel Dis, 13*(12), 1561-1566.

Plevy, S., Silverberg, M. S., Lockton, S., Stockfisch, T., Croner, L., Stachelski, J., Brown, M., Triggs, C., Chuang, E., Princen, F., & Singh, S. (2013). Combined serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients. *Inflamm Bowel Dis, 19*(6), 1139-1148.

Prideaux, L., De Cruz, P., Ng, S. C., & Kamm, M. A. (2012). Serological antibodies in inflammatory bowel disease: a systematic review. *Inflamm Bowel Dis, 18*(7), 1340-1355.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D.,

& Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59-65.

Rabilloud, T., Chevallet, M., Luche, S., & Lelong, C. (2010). Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics, 73*(11), 2064-2077.

Raine, T. (2014). Vedolizumab for inflammatory bowel disease: Changing the game, or more of the same? *United European Gastroenterology Journal, 2*(5), 333-344.

Rakonjac, J., Jovanovic, G., & Model, P. (1997). Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3. *Gene, 198*(1-2), 99-103.

Ravindran, R., Krishnan, V. V., Dhawan, R., Wunderlich, M. L., Lerche, N. W., Flynn, J. L., Luciw, P. A., & Khan, I. H. (2014). Plasma antibody profiles in non-human primate tuberculosis. *J Med Primatol, 43*(2), 59-71.

Reese, G. E., Constantinides, V. A., Simillis, C., Darzi, A. W., Orchard, T. R., Fazio, V. W., & Tekkis, P. P. (2006). Diagnostic precision of anti-Saccharomyces cerevisiae antibodies and perinuclear antineutrophil cytoplasmic antibodies in inflammatory bowel disease. *Am J Gastroenterol, 101*(10), 2410-2422.

Rhyner, C., Weichel, M., Flückiger, S., Hemmann, S., Kleber-Janke, T., & Crameri, R. (2004). Cloning allergens via phage display. *Methods, 32*(3), 212-218.

Robinson, W. H., DiGennaro, C., Hueber, W., Haab, B. B., Kamachi, M., Dean, E. J., Fournel, S., Fong, D., Genovese, M. C., de Vegvar, H. E., Skriner, K., Hirschberg, D. L., Morris, R. I., Muller, S., Pruijn, G. J., van Venrooij, W. J., Smolen, J. S., Brown, P. O., Steinman, L., & Utz, P. J. (2002). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med, 8*(3), 295-301.

Rocchi, A., Benchimol, E. I., Bernstein, C. N., Bitton, A., Feagan, B., Panaccione, R., Glasgow, K. W., Fernandes, A., & Ghosh, S. (2012). Inflammatory bowel disease: a Canadian burden of illness review. *Can J Gastroenterol, 26*(11), 811-817.

Roggenbuck, D., Hausdorf, G., Martinez-Gamboa, L., Reinhold, D., Büttner, T., Jungblut, P. R., Porstmann, T., Laass, M. W., Henker, J., Büning, C., Feist, E., & Conrad, K. (2009). Identification of GP2, the major zymogen granule membrane glycoprotein, as the autoantigen of pancreatic antibodies in Crohn's disease. *Gut, 58*(12), 1620-1628.

Rondot, S., Koch, J., Breitling, F., & Dübel, S. (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnology, 19*(1), 75-78.

Rossen, N. G., MacDonald, J. K., de Vries, E. M., D'Haens, G. R., de Vos, W. M., Zoetendal, E. G., & Ponsioen, C. Y. (2015). Fecal microbiota transplantation as novel therapy in gastroenterology: A systematic review. *World J Gastroenterol, 21*(17), 5359-5371.

Roulis, M., Bongers, G., Armaka, M., Salviano, T., He, Z., Singh, A., Seidler, U., Becker, C., Demengeot, J., Furtado, G. C., Lira, S. A., & Kollias, G. (2015). Host and microbiota interactions are critical for development of murine Crohn's-like ileitis. *Mucosal Immunol*.

Rump, J. A., Schölmerich, J., Gross, V., Roth, M., Helfesrieder, R., Rautmann, A., Lüdemann, J., Gross, W. L., & Peter, H. H. (1990). A new type of perinuclear anti-neutrophil cytoplasmic antibody (p-ANCA) in active ulcerative colitis but not in Crohn's disease. *Immunobiology, 181*(4-5), 406-413.

Saxon, A., Shanahan, F., Landers, C., Ganz, T., & Targan, S. (1990). A distinct subset of antineutrophil cytoplasmic antibodies is associated with inflammatory bowel disease. *J Allergy Clin Immunol, 86*(2), 202-210.

Schaubeck, M., Clavel, T., Calasan, J., Lagkouvardos, I., Haange, S. B., Jehmlich, N., Basic, M., Dupont, A., Hornef, M., Bergen, M. V., Bleich, A., & Haller, D. (2015). Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut*.

Schirwitz, C., Loeffler, F. F., Felgenhauer, T., Stadler, V., Breitling, F., & Bischoff, F. R. (2012). Sensing immune responses with customized peptide microarrays. *Biointerphases, 7*(1-4), 47.

Seehaus, T., Breitling, F., Dübel, S., Klewinghaus, I., & Little, M. (1992). A vector for the removal of deletion mutants from antibody libraries. *Gene, 114*(2), 235-237.

Seibold, F. (2005). ASCA: genetic marker, predictor of disease, or marker of a response to an environmental antigen? *Gut, 54*(9), 1212-1213.

Sendid, B., Colombel, J. F., Jacquinot, P. M., Faille, C., Fruit, J., Cortot, A., Lucidarme, D., Camus, D., & Poulain, D. (1996). Specific antibody response to oligomannosidic epitopes in Crohn's disease. *Clin Diagn Lab Immunol, 3*(2), 219-226.

Seow, C. H., Stempak, J. M., Xu, W., Lan, H., Griffiths, A. M., Greenberg, G. R., Steinhart, A. H., Dotan, N., & Silverberg, M. S. (2009). Novel anti-glycan antibodies related to inflammatory bowel disease diagnosis and phenotype. *American Journal of Gastroenterology, 104*(6), 1426-1434.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science, 228*(4705), 1315-1317.

Soltes, G., Hust, M., Ng, K. K., Bansal, A., Field, J., Stewart, D. I., Dübel, S., Cha, S., & Wiersma, E. J. (2007). On the influence of vector design on antibody phage display. *J Biotechnol, 127*(4), 626-637.

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., & Sermon, K. (2006). Whole-genome multiple displacement amplification from single cells. *Nat. Protocols, 1*(4), 1965-1970.

Stöcker, W., Otte, M., Ulrich, S., Normann, D., Finkbeiner, H., Stocker, K., Jantschek, G., & Scriba, P. C. (1987). Autoimmunity to pancreatic juice in Crohn's disease. Results of an autoantibody screening in patients with chronic inflammatory bowel disease. *Scand J Gastroenterol Suppl, 139*, 41-52.

Stratmann, T., & Kang, A. S. (2005). Cognate peptide-receptor ligand mapping by directed phage display. *Proteome Sci, 3*, 7.

Sutton, C. L., Kim, J., Yamane, A., Dalwadi, H., Wei, B., Landers, C., Targan, S. R., & Braun, J. (2000). Identification of a novel bacterial sequence associated with Crohn's disease. *Gastroenterology, 119*(1), 23-31.

Sutton, C. L., Yang, H., Li, Z., Rotter, J. I., Targan, S. R., & Braun, J. (2000). Familial expression of anti-Saccharomyces cerevisiae mannan antibodies in affected and unaffected relatives of patients with Crohn's disease. *Gut, 46*(1), 58-63.

Tamboli, C. P., Doman, D. B., & Patel, A. (2011). Current and future role of biomarkers in Crohn's disease risk assessment and treatment. *Clin Exp Gastroenterol, 4*, 127-140.

Taupp, M., Mewis, K., & Hallam, S. J. (2011). The art and design of functional metagenomic screens. *Curr Opin Biotechnol, 22*(3), 465-472.

Török, H. P., Glas, J., Hollay, H. C., Gruber, R., Osthoff, M., Tonenchi, L., Brückl, C., Mussack, T., Folwaczny, M., & Folwaczny, C. (2005). Serum antibodies in first-degree relatives of patients with IBD: a marker of disease susceptibility? A follow-up pilot-study after 7 years. *Digestion, 72*(2-3), 119-123.

Triantafillidis, J. K., Merikas, E., & Georgopoulos, F. (2011). Current and emerging drugs for the treatment of inflammatory bowel disease. *Drug Design, Development and Therapy, 5*, 185-210.

Türeci, O., Sahin, U., & Pfreundschuh, M. (1997). Serological analysis of human tumor antigens: molecular definition and implications. *Mol Med Today, 3*(8), 342-349.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature, 457*(7228), 480-484.

Urquhart, B. L., Cordwell, S. J., & Humphery-Smith, I. (1998). Comparison of predicted and observed properties of proteins encoded in the genome of Mycobacterium tuberculosis H37Rv. *Biochem Biophys Res Commun, 253*(1), 70-79.

van den Berg, S., Bowden, M. G., Bosma, T., Buist, G., van Dijl, J. M., van Wamel, W. J., de Vogel, C. P., van Belkum, A., & Bakker-Woudenberg, I. A. (2011). A multiplex assay for the quantification of antibody responses in Staphylococcus aureus infections in mice. *J Immunol Methods, 365*(1-2), 142-148.

Vieira, J., & Messing, J. (1987). Production of single-stranded plasmid DNA. *Methods Enzymol, 153*, 3-11.

Wang, J., Linnenbrink, M., Künzel, S., Fernandes, R., Nadeau, M.-J., Rosenstiel, P., & Baines, J. F. (2014). Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proceedings of the National Academy of Sciences, 111*(26), E2703-E2710.

Wei, B., Dalwadi, H., Gordon, L. K., Landers, C., Bruckner, D., Targan, S. R., & Braun, J. (2001). Molecular cloning of a Bacteroides caccae TonB-linked outer membrane protein identified by an inflammatory bowel disease marker antibody. *Infect Immun, 69*(10), 6044-6054.

Wei, B., Huang, T., Dalwadi, H., Sutton, C. L., Bruckner, D., & Braun, J. (2002). Pseudomonas fluorescens encodes the Crohn's disease-associated I2 sequence and T-cell superantigen. *Infection and Immunity, 70*(12), 6567-6575.

Wirtz, S., & Neurath, M. F. (2007). Mouse models of inflammatory bowel disease. *Adv Drug Deliv Rev, 59*(11), 1073-1083.

Zenewicz, L. A., Yin, X., Wang, G., Elinav, E., Hao, L., Zhao, L., & Flavell, R. A. (2013). IL-22 deficiency alters colonic microbiota to be transmissible and colitogenic. *J Immunol, 190*(10), 5306-5312.

Zhang, C. G., Chromy, B. A., & McCutchen-Maloney, S. L. (2005). Host-pathogen interactions: a proteomic view. *Expert Rev Proteomics, 2*(2), 187-202.

Zhou, M., Meyer, T., Koch, S., Koch, J., von Briesen, H., Benito, J. M., Soriano, V., Haberl, A., Bickel, M., Dübel, S., Hust, M., & Dietrich, U. (2013). Identification of a new epitope for HIV-neutralizing antibodies in the gp41 membrane proximal external region by an Env-tailored phage display library. *Eur J Immunol, 43*(2), 499-509.

11 Supplemental Information

11.1 Supplementary tables

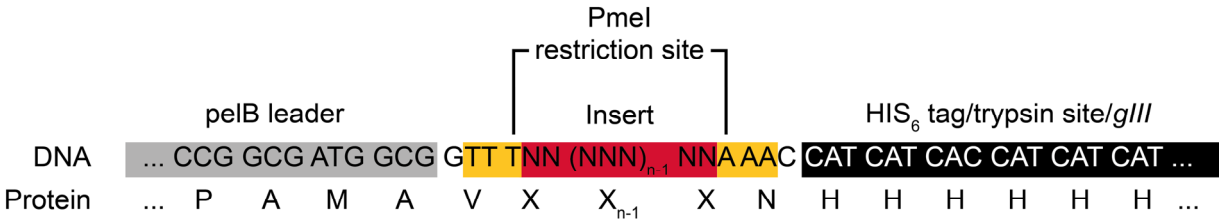
Suppl. Table 1: Summary of selected clones and heat map of reactivity with different IBD pools. Clones marked with asterisk were further analyzed. (S = single serum used for selection; 1-4 CD serum pools; 5-6 UC serum pools; 7 = healthy control serum pool)

Clone	Insert size (bp)	Copies	IBD serum							
			S	1	2	3	4	5	6	7
JOZ240-B11	172	1	1	0	0	0	0	0	0	0
JOZ240-B7	100	1	2	0	0	0	0	0	0	0
JOZ240-D12	34	1	3	0	0	0	0	0	0	0
JOZ240-D7*	181	1	1	0	1	0	0	0	1	0
JOZ240-D9	74	1	1	0	1	0	0	0	0	0
JOZ240-E3	178	2	2	0	0	0	0	0	0	0
JOZ240-E4	115	1	0	0	0	0	0	0	0	0
JOZ240-G3	199	1	2	0	1	0	1	0	1	1
JOZ240-G4	409	1	3	3	3	3	3	3	3	3
JOZ240-G5	280	1	2	0	1	0	0	0	0	0
JOZ240-H2*	121	1	2	0	0	0	0	0	0	0
JOZ241-A4*	229	1	1	0	1	0	2	0	1	0
JOZ241-C4*	349	1	2	0	1	0	1	0	0	0
JOZ241-E1	490	5	3	3	3	3	3	3	3	3
JOZ241-F4	94	1	1	1	1	0	1	0	0	1
JOZ241-F8	94	3	1	0	1	0	0	0	0	0
JOZ242-B2	163	1	0	0	1	0	0	0	0	1
JOZ242-C1	157	1	0	0	1	0	0	0	0	0
JOZ242-C11	277	1	1	1	0	0	0	0	0	0
JOZ242-C9*	328	1	2	1	1	0	1	0	0	0
JOZ242-G4*	217	1	1	2	0	0	1	1	1	0
JOZ242-H4	76	1	2	0	0	0	0	0	1	0
JOZ243-B9	118	1	2	0	1	0	0	0	0	0
JOZ243-F12*	31	3	3	1	0	0	2	0	2	0
JOZ243-C7	280	1	2	0	0	0	0	0	0	0
JOZ243-E11	181	1	2	1	0	0	0	0	0	0
JOZ243-E9*	208	1	2	1	1	0	2	1	1	0
JOZ243-F1	115	1	0	1	0	0	0	0	0	0
JOZ243-F11	196	1	1	0	0	0	1	0	0	0
JOZ243-F9*	296	1	1	0	1	1	1	0	0	0
JOZ243-G3	277	1	2	0	1	0	0	0	0	0

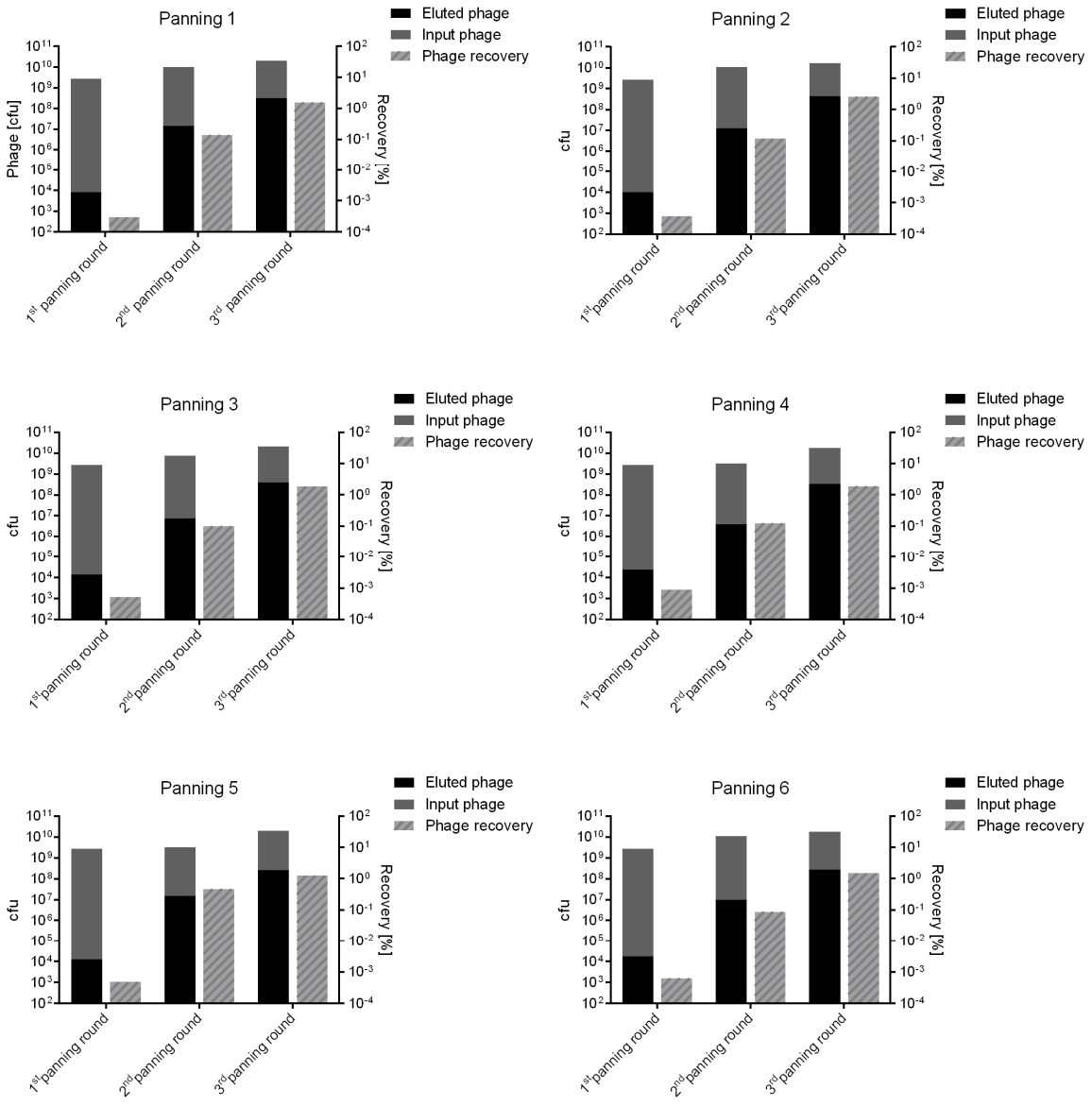
Suppl. Table 2: Summary of PCR amplified and cloned full-length genes

Full length protein	GenBank Accession	Identity DNA level (%)	Identity protein level (%)	Identity selected and amplified sequence (DNA level) (%)
JOZ191-G8	ALY14458.1	97.0	98.7	93
JOZ241-A4	AJE06446.1	99.1	98.9	100
JOZ241-C4	AAO81020.1	99.2	98.9	100
JOZ242-C9	ALE13104.1	99.1	99.8	100
JOZ243-F9	AII76920.1	-	-	-

11.2 Supplementary figures



Suppl. Figure 1: Illustration of the cloning site in the ORFeome phage display vector pHORF3. Only insert with 3n+1 nucleotides result in open reading frames with pelB leader sequence, hexahistidine tag and M13 minor coat protein III gene (*gIII*).



Suppl. Figure 2: Enrichment of binding oligopeptide phage in the unmatched panning approach. Enrichment in second panning round by more than two orders of magnitude and in third round by only about one order of magnitude. Phage titers were determined by infecting *E. coli* with dilution series of phage and spotting on ampicillin supplemented medium.

11.3 Sequences

11.3.1 Mouse approach

>JOZ156-H5
GGGACTTCATGGGGAACTTTTGGAATTTTAATTCCTATCGCGACTTCGATTTTCCCCG

>JOZ158-E11
TCTGGCCTGCTGACGGATATGAACCGGGACACGGTCAGCCTTCCTTTGACAAGCAGTTTGCCCGCGACTGGCTGAAAG
AAAATGACGGTCATGACTGGACTCTTCCTCAGGAGATCGTTGA

>JOZ158-G8
CAAACGACCTGAACGCGCTGCTTGAGACAAACATATCCGCGGGAACCAGAACTACCCAGGTATCTATGAGCTATTTCG
GGGAAATGCTTATGAGCCATATCGCGGACAATGAATTCTTAAGCGGCTCGGAAGATGAAAAGGCGGCTCACGCAAATG
CGCTTGC

11.3.2 Human approach

11.4 Sequence alignments

11.4.1 JOZ191-G8

11.4.1.1 DNA alignment

```
                                1                                        40
           ALY14458.1      (1)  ATGGTTAAATTACCGCCGCTGAGTCTCTACATTCACATCC
JOZ191-G8 full-length      (1)  ATGGTTAAATTACCGCCGCTGAGTCTCTACATTCACATCC
            JOZ191-G8      (1)  ----------------------------------------
                               41                                       80
           ALY14458.1     (41)  CGTGGTGCGTGCAGAAATGCCCGTACTGCGATTTCAACTC
JOZ191-G8 full-length     (41)  CGTGGTGCGTGCAGAAATGCCCGTACTGCGATTTCAACTC
            JOZ191-G8      (1)  ----------------------------------------
                               81                                      120
           ALY14458.1     (81)  TCACGCGTTGAAAGGAGAAGTGCCGCACGACGATTACGTT
JOZ191-G8 full-length     (81)  TCACGCGTTGAAAGGAGAAGTGCCGCACGACGATTATGTT
            JOZ191-G8      (1)  ----------------------------------------
                              121                                      160
           ALY14458.1    (121)  CAGCATCTGCTTTGCGATCTGGACAATGATGTGGCTTACG
JOZ191-G8 full-length    (121)  CAGCATCTGCTTAACGATCTGGACAACGATGTGGCTTACG
            JOZ191-G8      (1)  ----------------------------------------
                              161                                      200
           ALY14458.1    (161)  CTCAGGGCCGTGAAGTAAAGACAATCTTTATTGGCGGTGG
JOZ191-G8 full-length    (161)  CTCAGGGCCGTGAAGTAAAGACAATCTTTATTGGCGGTGG
            JOZ191-G8      (1)  ----------------------------------------
                              201                                      240
           ALY14458.1    (201)  TACGCCGAGCCTGCTTTCCGGCCCGGCGATGCAAACGCTG
JOZ191-G8 full-length    (201)  TACGCCGAGCCTGCTTTCCGGCCCGGCGATGCAAACGCTG
            JOZ191-G8      (1)  ----------------------------------------
                              241                                      280
           ALY14458.1    (241)  CTGGACGGCGTGCGTGCGCGTTTGCCGCTGGCAGCGGATG
JOZ191-G8 full-length    (241)  CTGGACGGCGTGCGTGCGCGTTTGCCGCTGGCAGCGGATG
            JOZ191-G8      (1)  ----------------------------------------
                              281                                      320
           ALY14458.1    (281)  CAGAAATTACTATGGAAGCGAACCCTGGTACGGTAGAAGC
JOZ191-G8 full-length    (281)  CAGAAATTACTATGGAAGCGAACCCTGGCACGGTAGAAGC
            JOZ191-G8      (1)  ----------------------------------------
```

```
                                 321                                 360
            ALY14458.1    (321) CGATCGCTTTGTCGATTATCAGCGTGCTGGTGTGAACCGC
JOZ191-G8 full-length     (321) CGATCGCTTTGTCGATTATCAGCGTGCTGGTGTGAACCGC
            JOZ191-G8       (1) -------------------------------------
                                 361                                 400
            ALY14458.1    (361) ATCTCTATTGGTGTGCAGAGTTTTAGCGAAGAAAAGCTGA
JOZ191-G8 full-length     (361) ATCTCTATTGGTGTGCAGAGTTTTAGCGAAGAAAAGCTGA
            JOZ191-G8       (1) -------------------------------------
                                 401                                 440
            ALY14458.1    (401) AACGACTTGGGCGTATTCATGGCCCGCAAGAAGCGAAACG
JOZ191-G8 full-length     (401) AACGACTTGGGCGTATTCATGGCCCGCAAGAAGCGAAACG
            JOZ191-G8       (1) -------------------------------------
                                 441                                 480
            ALY14458.1    (441) CGCGGCAAAGCTGGCGAGCGGGCTGGGGCTGCGTAGCTTT
JOZ191-G8 full-length     (441) CGCGGCGAAGCTGGCGAGCGGTTTAGGGTTACGTAGCTTT
            JOZ191-G8       (1) -------------------------------------
                                 481                                 520
            ALY14458.1    (481) AACCTCGATTTGATGCATGGACTACCGGATCAATCACTGG
JOZ191-G8 full-length     (481) AACCTTGATTTGATGCATGGGCTGCCGGATCAATCACTGG
            JOZ191-G8       (1) -------------------------------------
                                 521                                 560
            ALY14458.1    (521) AAGAGGCGCTTGGCGATCTACGCCAGGCCATTGAACTGAA
JOZ191-G8 full-length     (521) AAGAGGCGCTTGGCGATCTACGCCAGGCCATTGAACTGAA
            JOZ191-G8       (1) -------------------------------------
                                 561                                 600
            ALY14458.1    (561) TCCGCCGCATCTTTCCTGGTATCAACTGACCATCGAACCT
JOZ191-G8 full-length     (561) TCCGCCGCATCTTTCCTGGTATCAACTGACCATCGAACCC
            JOZ191-G8       (1) ----------------------------CCATCGAACCT
                                 601                                 640
            ALY14458.1    (601) AATACGCTGTTTGGTTCACGCCCGCCGGTACTGCCGGACG
JOZ191-G8 full-length     (601) AATACGCTGTTTGGTTCGCGACCACCGGTGCTGCCGGACG
            JOZ191-G8      (12) AATACGCTGTTTGGTTCACGCCCGCCGGTACTGCCGGACG
                                 641                                 680
            ALY14458.1    (641) ACGACGCGCTGTGGGATATATTCGAACAGGGGCATCAGTT
JOZ191-G8 full-length     (641) ATGACGCGTTGTGGGATATATTCGAACAGGGGCATCAGTT
            JOZ191-G8      (52) ACGACGCGCTGTGGGATATATTCGAACAGGGGCATCAGTT
                                 681                                 720
            ALY14458.1    (681) ATTAACCGCAGCGGGTTATCAGCAATATGAAACTTCCGCT
JOZ191-G8 full-length     (681) ATTAACCGCAGCGGGTTATCAGCAATATGAAACTTCCGCT
            JOZ191-G8      (92) ATTAACCGC-----------------------------
                                 721                                 760
            ALY14458.1    (721) TACGCCAAACCCAGTTATCAGTGCCAGCACAATCTCAACT
JOZ191-G8 full-length     (721) TACGCCAAACCCGGTTATCAGTGCCAGCACAATCTCAACT
            JOZ191-G8     (101) -------------------------------------
                                 761                                 800
            ALY14458.1    (761) ACTGGCGCTTTGGCGACTACATCGGTATTGGCTGCGGCGC
JOZ191-G8 full-length     (761) ACTGGCGCTTTGGTGACTACATCGGTATTGGCTGCGGCGC
            JOZ191-G8     (101) -------------------------------------
                                 801                                 840
            ALY14458.1    (801) GCACGGCAAAGTGACCTTCCCGGATGGGCGCATTCTGCGT
JOZ191-G8 full-length     (801) ACACGGCAAAGTGACCTTCCCGGATGGGCGCATTCTGCGT
            JOZ191-G8     (101) -------------------------------------
                                 841                                 880
            ALY14458.1    (841) ACCACTAAAACGCGTCATCCGCGTGGTTTTATGCAAGGAA
JOZ191-G8 full-length     (841) ACCACCAAAACGCGTCATCCGCGTGGTTTTATGCAAGGAA
            JOZ191-G8     (101) -------------------------------------
```

```
                                881                                       920
           ALY14458.1    (881) GGTATCTGGAAAGCCAGCGTGATGTCGAAGCCGCAGATAA
JOZ191-G8 full-length    (881) GGTATCTGGAAAGCCAGCGTGATGTCGAAGCCACAGATAA
            JOZ191-G8    (101) ----------------------------------------
                                921                                       960
           ALY14458.1    (921) GCCGTTTGAGTTCTTTATGAACCGCTTTCGCTTGCTGGAG
JOZ191-G8 full-length    (921) GCCGTTTGAGTTCTTTATGAATCGCTTCCGTCTGCTGGAG
            JOZ191-G8    (101) ----------------------------------------
                                961                                      1000
           ALY14458.1    (961) GCCGCGCCACGCGCAGAGTTTAGTGCGTATACCGGGCTTT
JOZ191-G8 full-length    (961) GCCGCGCCGCGCGTGGAGTTTATTGCGTATACCGGGCTTT
            JOZ191-G8    (101) ----------------------------------------
                               1001                                      1040
           ALY14458.1   (1001) GCGAAGATGTGATTCGCCCACAGTTAGACGAGGCGATTGC
JOZ191-G8 full-length   (1001) GCGAAGATGTGATTCGCCCACAGTTAGACGAGGCGATTGC
            JOZ191-G8    (101) ----------------------------------------
                               1041                                      1080
           ALY14458.1   (1041) CCAGGGTTATCTCACCGAATGTGCGGATTACTGGCAGATA
JOZ191-G8 full-length   (1041) CCAGGGTTATCTCACCGAATGTGCGGATTACTGGCAGATA
            JOZ191-G8    (101) ----------------------------------------
                               1081                                      1120
           ALY14458.1   (1081) ACGGAACATGGGAAGCTGTTTTTAAATTCGCTGCTGGAGC
JOZ191-G8 full-length   (1081) ACGGAACATGGGAAGCTGTTTTTAAATTCGCTGCTGGAGC
            JOZ191-G8    (101) ----------------------------------------
                               1121                                      1160
           ALY14458.1   (1121) TTTTTCTGGCTGAG--------------------------
JOZ191-G8 full-length   (1121) TTTTTCTGGCTGAGGCGGCCGCACTCGAGCACCACCACCA
            JOZ191-G8    (101) ----------------------------------------
                               1161
           ALY14458.1   (1135) ----------
JOZ191-G8 full-length   (1161) CCACCACTGA
            JOZ191-G8    (101) ----------
```

11.4.1.2 Protein alignment

```
                                  1                                        40
           ALY14458.1      (1) MVKLPPLSLYIHIPWCVQKCPYCDFNSHALKGEVPHDDYV
JOZ191-G8 full-length      (1) MVKLPPLSLYIHIPWCVQKCPYCDFNSHALKGEVPHDDYV
            JOZ191-G8      (1) ----------------------------------------
                                 41                                        80
           ALY14458.1     (41) QHLLCDLDNDVAYAQGREVKTIFIGGGTPSLLSGPAMQTL
JOZ191-G8 full-length     (41) QHLLNDLDNDVAYAQGREVKTIFIGGGTPSLLSGPAMQTL
            JOZ191-G8      (1) ----------------------------------------
                                 81                                       120
           ALY14458.1     (81) LDGVRARLPLAADAEITMEANPGTVEADRFVDYQRAGVNR
JOZ191-G8 full-length     (81) LDGVRARLPLAADAEITMEANPGTVEADRFVDYQRAGVNR
            JOZ191-G8      (1) ----------------------------------------
                                121                                       160
           ALY14458.1    (121) ISIGVQSFSEEKLKRLGRIHGPQEAKRAAKLASGLGLRSF
JOZ191-G8 full-length    (121) ISIGVQSFSEEKLKRLGRIHGPQEAKRAAKLASGLGLRSF
            JOZ191-G8      (1) ----------------------------------------
                                161                                       200
           ALY14458.1    (161) NLDLMHGLPDQSLEEALGDLRQAIELNPPHLSWYQLTIEP
JOZ191-G8 full-length    (161) NLDLMHGLPDQSLEEALGDLRQAIELNPPHLSWYQLTIEP
            JOZ191-G8      (1) ----------------------------------------IEP
```

```
                              201                                       240
           ALY14458.1   (201) NTLFGSRPPVLPDDDALWDIFEQGHQLLTAAGYQQYETSA
JOZ191-G8 full-length   (201) NTLFGSRPPVLPDDDALWDIFEQGHQLLTAAGYQQYETSA
            JOZ191-G8     (4) NTLFGSRPPVLPDDDALWDIFEQGHQLLT-----------
                              241                                       280
           ALY14458.1   (241) YAKPSYQCQHNLNYWRFGDYIGIGCGAHGKVTFPDGRILR
JOZ191-G8 full-length   (241) YAKPGYQCQHNLNYWRFGDYIGIGCGAHGKVTFPDGRILR
            JOZ191-G8    (33) ------------------------------------------
                              281                                       320
           ALY14458.1   (281) TTKTRHPRGFMQGRYLESQRDVEAADKPFEFFMNRFRLLE
JOZ191-G8 full-length   (281) TTKTRHPRGFMQGRYLESQRDVEATDKPFEFFMNRFRLLE
            JOZ191-G8    (33) ------------------------------------------
                              321                                       360
           ALY14458.1   (321) AAPRAEFSAYTGLCEDVIRPQLDEAIAQGYLTECADYWQI
JOZ191-G8 full-length   (321) AAPRVEFIAYTGLCEDVIRPQLDEAIAQGYLTECADYWQI
            JOZ191-G8    (33) ------------------------------------------
                              361                           389
           ALY14458.1   (361) TEHGKLFLNSLLELFLAE-----------
JOZ191-G8 full-length   (361) TEHGKLFLNSLLELFLAEAAALEHHHHHH
            JOZ191-G8    (33) ----------------------------
```

11.4.2 JOZ241-A4

11.4.2.1 DNA alignment

```
                              1                                         40
           AJE06446.1     (1) ATGGCAACTGAGAACGATGATTCCCAGTCTATGAACGGTA
JOZ241-A4 full-length     (1) ATGGCAACTGAGAACGATGATTCCCAGTCTATGAACGGTA
            JOZ241-A4     (1) ---------------ATGATTCCCAGTCTATGAACGGTA
                              41                                        80
           AJE06446.1    (41) TTCTGTCCAATATATCTCGGGATGTGGTGGACACCAGCTG
JOZ241-A4 full-length    (41) TTTTGTCCAACATATCTCGGGATGTGGTGGACACCAGCTG
            JOZ241-A4    (25) TTTTGTCCAACATATCTCGGGATGTGGTGGACACCAGCTG
                              81                                       120
           AJE06446.1    (81) GGATATTCCTCCGCTGACCTTCCCCAAATTGCAGCCTACC
JOZ241-A4 full-length    (81) GGATATTCCTCCGCTGACCTTCCCCAAATTGCAGCCTACC
            JOZ241-A4    (65) GGATATTCCTCCGCTGACCTTCCCCAAATTGCAGCCTACC
                             121                                       160
           AJE06446.1   (121) GATTCGGAAGAAGCCGATGGTGCGGATGAGGCTTCGCTTG
JOZ241-A4 full-length   (121) GATTCGGAAGAAGCCGATGGTGCGGATGAGGCTTCGCTTG
            JOZ241-A4   (105) GATTCGGAAGAAGCCGATGGTGCGGATGAGGCTTCGCTTG
                             161                                       200
           AJE06446.1   (161) CTGAGCCCCGCACCATCGCTTCCTCCGTGCCTGCGCAAAC
JOZ241-A4 full-length   (161) CTGAGCCCCGCACCATCGCTTCCTCCGTGCCTGCGCAAAC
            JOZ241-A4   (145) CTGAGCCCCGCACCATCGCTTCCTCCGTGCCTGCGCAAAC
                             201                                       240
           AJE06446.1   (201) GGAACAGCCGGTTTCGGCTGCGCCAACAATCGCGGATATT
JOZ241-A4 full-length   (201) GGAACAGCCGGTTTCGGCTGCGCCAACAATCGCGGATATT
            JOZ241-A4   (185) GGAACAGCCGGTTTCGGCTGCGCCAACAATCGCGGATATT
                             241                                       280
           AJE06446.1   (241) CCCAGTATGCCGCTCGATACCGTCGATTTCGATGATGCGG
JOZ241-A4 full-length   (241) CCCAGTATGCCGCTCGACACCATCGATTTCGATGATGCGG
            JOZ241-A4   (225) CCCAG-------------------------------------
```

```
                                 281                                 320
            AJE06446.1    (281)  TTGCACCAGCCGATGCCGAGGAAACCGCCACTTTGCCACC
JOZ241-A4 full-length     (281)  TTGCACCAGCCGATGCCGAGGAAACCGCCACTTTGCCACC
             JOZ241-A4    (230)  ----------------------------------------
                                 321                                 360
            AJE06446.1    (321)  CCTCGTCTCTTCGACAGATGAACAGGACGTCGAAGCGACT
JOZ241-A4 full-length     (321)  CCTCGTCTCTTCGACAGATGAACAGGACGCCGAAGCGACT
             JOZ241-A4    (230)  ----------------------------------------
                                 361                                 400
            AJE06446.1    (361)  CAGGCCTTTGATACATCAGGTGGCATAACCGACCTAGAGG
JOZ241-A4 full-length     (361)  CAGGCCTTTGACACTTCGGATGGCATAACCGACCTAGAGG
             JOZ241-A4    (230)  ----------------------------------------
                                 401                                 440
            AJE06446.1    (401)  AGCTGGCCGCGGAAGATCCGGAACATGATCCGCTGAAAAC
JOZ241-A4 full-length     (401)  AACTTGCCGCGGAAGATCCGGAACATGATCCGCTGAAAAC
             JOZ241-A4    (230)  ----------------------------------------
                                 441                                 480
            AJE06446.1    (441)  CATTCCGGTCGCACCGAAAGAGCCGTCTGCGGACAAGCGA
JOZ241-A4 full-length     (441)  CATTCCGGTCGCACCGAAAGAGCCGTCTGCGGACAAGCGA
             JOZ241-A4    (230)  ----------------------------------------
                                 481                                 520
            AJE06446.1    (481)  AAGATTATTGCCATTGTCGCTGCGATTGCGGCAGTGGCGG
JOZ241-A4 full-length     (481)  AAGATTATTGCCATTGTCGCTGCGATTGCGGCAGTGGCGG
             JOZ241-A4    (230)  ----------------------------------------
                                 521                                 560
            AJE06446.1    (521)  TTGCTATTGCCGGTGTCATCGTGGTACGCAATCGTATCAA
JOZ241-A4 full-length     (521)  TTGCTATTGCCGGTGTCATCGTGGTACGCAATCGTATCAA
             JOZ241-A4    (230)  ----------------------------------------
                                 561                                 600
            AJE06446.1    (561)  TTCGAAAGAACAGCGTGCGGCAGTGGTGGTATGTGAGAAG
JOZ241-A4 full-length     (561)  TTCGAAAGAACAGCGTGCGGCAGTGGTGGTATGTGAGAAG
             JOZ241-A4    (230)  ----------------------------------------
                                 601                                 640
            AJE06446.1    (601)  GCGAAATCCAAATACTCCAAAGCCAACGAAAAACTGCAAA
JOZ241-A4 full-length     (601)  GCGAAATCCAAATACTCCAAAGCCAACGAAAAACTGCAAA
             JOZ241-A4    (230)  ----------------------------------------
                                 641                                 680
            AJE06446.1    (641)  ACGCCATCGAACAGGCAACTAGTCTGCAGGGCACTGCGGC
JOZ241-A4 full-length     (641)  ACGCCATCGAACAGGCAACTAGTCTGCAGGGCACTGCGGC
             JOZ241-A4    (230)  ----------------------------------------
                                 681                                 720
            AJE06446.1    (681)  AAACCAGGTGGCGGATGCCGCCACAATCGACCAGCTCACT
JOZ241-A4 full-length     (681)  AAACCAGGTGGCGGATGCCGCCACAATCGACCAGCTCACT
             JOZ241-A4    (230)  ----------------------------------------
                                 721                                 760
            AJE06446.1    (721)  CAGGCCGTGACCAAGGCGCAAAACCTCAAGACGGTCGGCG
JOZ241-A4 full-length     (721)  CAGGCCGTGACCAAGGCGCAAAACCTCAAGACGGTCGGCG
             JOZ241-A4    (230)  ----------------------------------------
                                 761                                 800
            AJE06446.1    (761)  GATGTTCGGTGCTGCTGTCCGAAACCACACTGCGTGCGCA
JOZ241-A4 full-length     (761)  GATGTTCGGTGCTGCTGTCCGAAACCACACTGCGTGCGCA
             JOZ241-A4    (230)  ----------------------------------------
                                 801                                 840
            AJE06446.1    (801)  CGCCCGTAGCATGAGCAAGCAGATCAGCAATATCAAAGAA
JOZ241-A4 full-length     (801)  CGCCCGTAGCATGAGCAAGCAGATCAGCAATATCAAAGAA
             JOZ241-A4    (230)  ----------------------------------------
```

```
                                841                                 880
           AJE06446.1    (841) CAGACCAAGGCCGTTACCGCAGCAGCCACAGCTGTGGACG
JOZ241-A4 full-length    (841) CAGACCAAGGCCGTTACCGCAGCAGCCACAGCTGTGGACG
            JOZ241-A4    (230) ----------------------------------------
                                881                                 920
           AJE06446.1    (881) CCAGCAAGCGAGCTTTGGAAGTGAATAAGACGACTGCCGC
JOZ241-A4 full-length    (881) CCAGCAAGCGAGCTTTGGAAGTGAATAAGACGACTGCCGC
            JOZ241-A4    (230) ----------------------------------------
                                921                                 960
           AJE06446.1    (921) ATTGCAGAAGGCCATTGCCGACGCAAAGACGTTGCTTGAC
JOZ241-A4 full-length    (921) ATTGCAGAAGGCCATTGCCGACGCAAAGACGTTGCTTGAC
            JOZ241-A4    (230) ----------------------------------------
                                961                                1000
           AJE06446.1    (961) GGTTCTGCAGGCCAGGTTGCGGACGAGTCGACACGTGACG
JOZ241-A4 full-length    (961) GGTTCTGCAGGCCAGGTTGCGGACGAGTCGACACGTGACG
            JOZ241-A4    (230) ----------------------------------------
                               1001                                1040
           AJE06446.1   (1001) CGCTTGCCAAGGCGATTGACGCTGCCCAGAAACTCCTCGA
JOZ241-A4 full-length   (1001) CGCTTGCCAAGGCGATTGACGCTGCCCAGAAACTCCTCGA
            JOZ241-A4    (230) ----------------------------------------
                               1041                                1080
           AJE06446.1   (1041) TGGCAAAAGCACCGATGTTGCGGCCATGCAGAACGCTTCG
JOZ241-A4 full-length   (1041) TGGCAAAAGCACCGATGTTGCGGCCATGCAGAACGCTTCG
            JOZ241-A4    (230) ----------------------------------------
                               1081                                1120
           AJE06446.1   (1081) AAGGTGATTTCCTCGACTTCTGACGGCGTAAACAAGTCTG
JOZ241-A4 full-length   (1081) AAGGTGATTTCCTCGACTTCTGACGGCGTAAACAAGTCTG
            JOZ241-A4    (230) ----------------------------------------
                               1121                                1160
           AJE06446.1   (1121) TGGCGGAGCAAACTGCGGCCAATGCCCAGTCGAATACCAA
JOZ241-A4 full-length   (1121) TGGCGGAGCAAACTGCGGCCAATGCCCAGTCGAATACCAA
            JOZ241-A4    (230) ----------------------------------------
                               1161                                1200
           AJE06446.1   (1161) TTCGAATACCAATAACGGCACTACGAACCGGCGTTACACA
JOZ241-A4 full-length   (1161) TTCGAATACCAATAACGGCACTACGAACCGGCGTTACACA
            JOZ241-A4    (230) ----------------------------------------
                               1201                                1240
           AJE06446.1   (1201) TATACGTGGCAATTTGGTCCGGGCTCTGGAAACGGTGTGA
JOZ241-A4 full-length   (1201) TATACGTGGCAATTTGGTCCGGGCTCTGGAAACGGTGTGA
            JOZ241-A4    (230) ----------------------------------------
                               1241                                1280
           AJE06446.1   (1241) ACGGTGGTGGTTCGACCGGTGGCAATAACAATGGTGGCTC
JOZ241-A4 full-length   (1241) ACGGTGGTGGTTCGACCGGTGGCAATAACAATGGTGGCTC
            JOZ241-A4    (230) ----------------------------------------
                               1281                                1320
           AJE06446.1   (1281) AACCGGTGGTAATGGCGATTCGACCGGCAACGGCAACGGC
JOZ241-A4 full-length   (1281) AACCGGTGGTAATGGCGATTCGACCGGCAACGGCAACGGC
            JOZ241-A4    (230) ----------------------------------------
                               1321                                1360
           AJE06446.1   (1321) GATTCAAACGGCGGCGGTAACGGCGATTCGAATGACGGTA
JOZ241-A4 full-length   (1321) GATTCAAACGGCGGCGGTAACGGCGATTCGAATGACGGTA
            JOZ241-A4    (230) ----------------------------------------
                               1361                                1400
           AJE06446.1   (1361) ATGGTGGTTCGACCGGCAACGGCAACGGCGATTCAAACGG
JOZ241-A4 full-length   (1361) ATGGCGATTCGACCGGCAACGGCAACGGCGATTCAAACGG
            JOZ241-A4    (230) ----------------------------------------
```

```
                                1401                                    1440
            AJE06446.1  (1401) CGGCGGTAACGGCGATTCGAATGACGGTAATGGTGGTTCG
JOZ241-A4 full-length  (1401) CGGCGGTAACGGCGATTCGAATGACGGTAATGGTGGTTCG
             JOZ241-A4   (230) ----------------------------------------
                                1441                                    1480
            AJE06446.1  (1441) ACCGGCAACGGCAACGGCGATTCAGATACAGGTACTGCTG
JOZ241-A4 full-length  (1441) ACCGGCAACGGCAACGGCGATTCAGATACAGGTACTGCTG
             JOZ241-A4   (230) ----------------------------------------
                                1481                                    1520
            AJE06446.1  (1481) GCGAA-----------------------------------
JOZ241-A4 full-length  (1481) GCGAAGCGGCCGCACTCGAGCACCACCACCACCACCACTG
             JOZ241-A4   (230) ----------------------------------------
                                1521
            AJE06446.1  (1486) -
JOZ241-A4 full-length  (1521) A
             JOZ241-A4   (230) -
```

11.4.2.2 Protein alignment

```
                                1                                       40
            AJE06446.1     (1) MATENDDSQSMNGILSNISRDVVDTSWDIPPLTFPKLQPT
JOZ241-A4 full-length      (1) MATENDDSQSMNGILSNISRDVVDTSWDIPPLTFPKLQPT
             JOZ241-A4     (1) ------DSQSMNGILSNISRDVVDTSWDIPPLTFPKLQPT
                                41                                      80
            AJE06446.1    (41) DSEEADGADEASLAEPRTIASSVPAQTEQPVSAAPTIADI
JOZ241-A4 full-length     (41) DSEEADGADEASLAEPRTIASSVPAQTEQPVSAAPTIADI
             JOZ241-A4    (35) DSEEADGADEASLAEPRTIASSVPAQTEQPVSAAPTIADI
                                81                                     120
            AJE06446.1    (81) PSMPLDTVDFDDAVAPADAEETATLPPLVSSTDEQDVEAT
JOZ241-A4 full-length     (81) PSMPLDTIDFDDAVAPADAEETATLPPLVSSTDEQDAEAT
             JOZ241-A4    (75) P---------------------------------------
                               121                                     160
            AJE06446.1   (121) QAFDTSGGITDLEELAAEDPEHDPLKTIPVAPKEPSADKR
JOZ241-A4 full-length    (121) QAFDTSDGITDLEELAAEDPEHDPLKTIPVAPKEPSADKR
             JOZ241-A4    (76) ----------------------------------------
                               161                                     200
            AJE06446.1   (161) KIIAIVAAIAAVAVAIAGVIVVRNRINSKEQRAAVVVCEK
JOZ241-A4 full-length    (161) KIIAIVAAIAAVAVAIAGVIVVRNRINSKEQRAAVVVCEK
             JOZ241-A4    (76) ----------------------------------------
                               201                                     240
            AJE06446.1   (201) AKSKYSKANEKLQNAIEQATSLQGTAANQVADAATIDQLT
JOZ241-A4 full-length    (201) AKSKYSKANEKLQNAIEQATSLQGTAANQVADAATIDQLT
             JOZ241-A4    (76) ----------------------------------------
                               241                                     280
            AJE06446.1   (241) QAVTKAQNLKTVGGCSVLLSETTLRAHARSMSKQISNIKE
JOZ241-A4 full-length    (241) QAVTKAQNLKTVGGCSVLLSETTLRAHARSMSKQISNIKE
             JOZ241-A4    (76) ----------------------------------------
                               281                                     320
            AJE06446.1   (281) QTKAVTAAATAVDASKRALEVNKTTAALQKAIADAKTLLD
JOZ241-A4 full-length    (281) QTKAVTAAATAVDASKRALEVNKTTAALQKAIADAKTLLD
             JOZ241-A4    (76) ----------------------------------------
                               321                                     360
            AJE06446.1   (321) GSAGQVADESTRDALAKAIDAAQKLLDGKSTDVAAMQNAS
JOZ241-A4 full-length    (321) GSAGQVADESTRDALAKAIDAAQKLLDGKSTDVAAMQNAS
             JOZ241-A4    (76) ----------------------------------------
```

```
                                361                                     400
            AJE06446.1   (361) KVISSTSDGVNKSVAEQTAANAQSNTNSNTNNGTTNRRYT
JOZ241-A4 full-length    (361) KVISSTSDGVNKSVAEQTAANAQSNTNSNTNNGTTNRRYT
             JOZ241-A4    (76) ----------------------------------------
                                401                                     440
            AJE06446.1   (401) YTWQFGPGSGNGVNGGGSTGGNNNGGSTGGNGDSTGNGNG
JOZ241-A4 full-length    (401) YTWQFGPGSGNGVNGGGSTGGNNNGGSTGGNGDSTGNGNG
             JOZ241-A4    (76) ----------------------------------------
                                441                                     480
            AJE06446.1   (441) DSNGGGNGDSNDGNGGSTGNGNGDSNGGGNGDSNDGNGGS
JOZ241-A4 full-length    (441) DSNGGGNGDSNDGNGDSTGNGNGDSNGGGNGDSNDGNGGS
             JOZ241-A4    (76) ----------------------------------------
                                481                         507
            AJE06446.1   (481) TGNGNGDSDTGTAGE------------
JOZ241-A4 full-length    (481) TGNGNGDSDTGTAGEAAALEHHHHHH-
             JOZ241-A4    (76) ---------------------------
```

11.4.3 JOZ241-C4

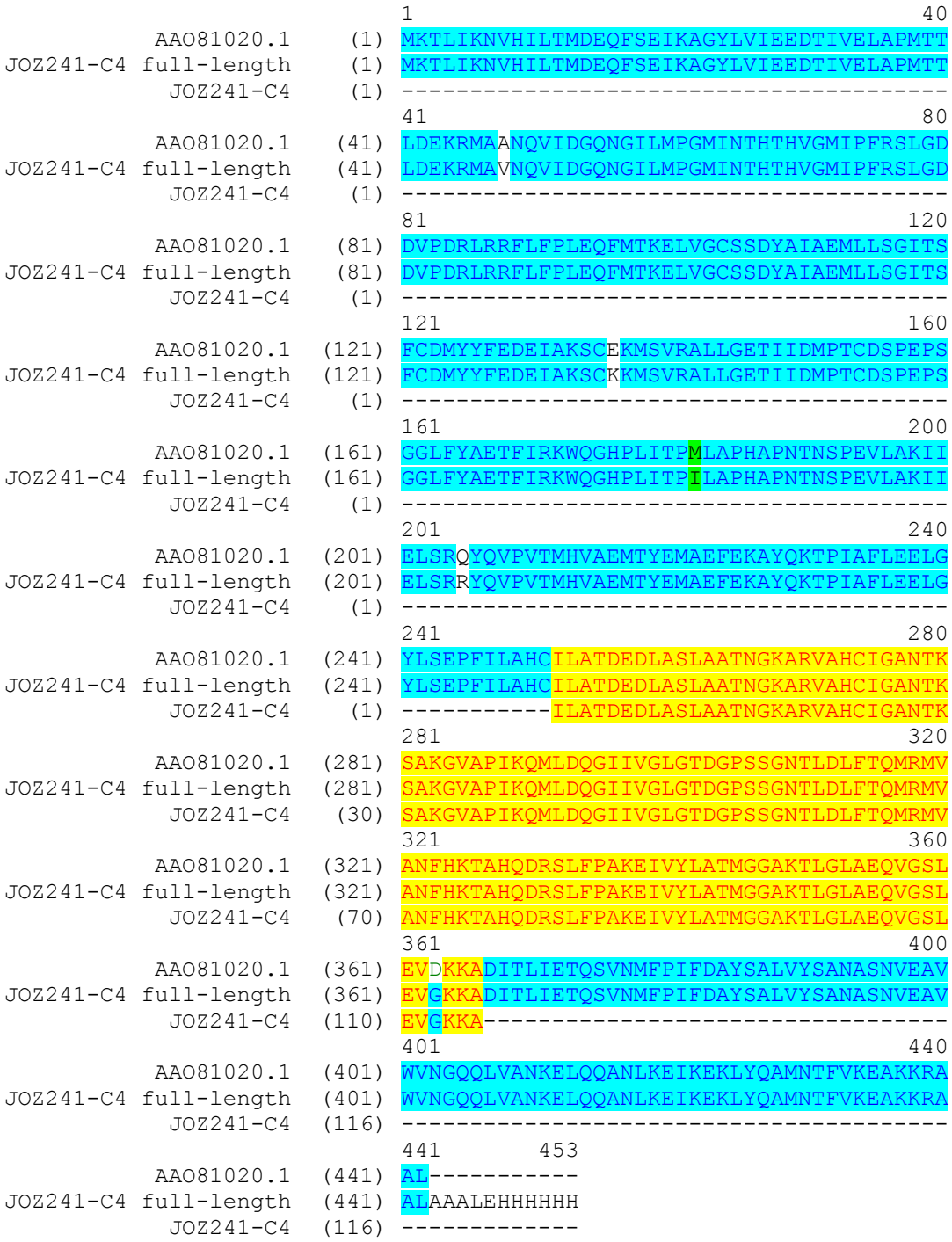
11.4.3.1 DNA alignment

```
                                  1                                     40
            AAO81020.1       (1) ATGAAAACACTCATTAAAAATGTGCATATACTAACAATGG
JOZ241-C4 full-length        (1) ATGAAAACACTCATTAAAAATGTGCATATACTAACAATGG
             JOZ241-C4       (1) ----------------------------------------
                                 41                                     80
            AAO81020.1      (41) ATGAGCAATTTTCAGAAATCAAAGCCGGCTATTTAGTGAT
JOZ241-C4 full-length       (41) ATGAGCAATTTTCAGAAATCAAAGCCGGCTATTTAGTGAT
             JOZ241-C4       (1) ----------------------------------------
                                 81                                    120
            AAO81020.1      (81) TGAAGAAGACACAATTGTGGAATTAGCACCAATGACCACA
JOZ241-C4 full-length       (81) TGAAGAAGATACAATTGTGGAATTAGCACCAATGACCACA
             JOZ241-C4       (1) ----------------------------------------
                                121                                    160
            AAO81020.1     (121) CTTGATGAAAAGCGAATGGCTGCAAATCAAGTAATCGATG
JOZ241-C4 full-length      (121) CTTGATGAAAAGCGAATGGCTGTAAATCAAGTAATCGATG
             JOZ241-C4       (1) ----------------------------------------
                                161                                    200
            AAO81020.1     (161) GTCAAAATGGAATTTTAATGCCTGGGATGATTAACACCCA
JOZ241-C4 full-length      (161) GTCAAAATGGAATTTTAATGCCTGGGATGATTAACACCCA
             JOZ241-C4       (1) ----------------------------------------
                                201                                    240
            AAO81020.1     (201) TACCCATGTAGGCATGATTCCGTTTCGTTCGTTAGGAGAC
JOZ241-C4 full-length      (201) TACCCATGTAGGCATGATTCCGTTTCGTTCGTTAGGAGAC
             JOZ241-C4       (1) ----------------------------------------
                                241                                    280
            AAO81020.1     (241) GATGTGCCAGATCGACTCCGGCGTTTTCTTTTTCCATTAG
JOZ241-C4 full-length      (241) GATGTGCCAGATCGACTCCGGCGTTTTCTTTTTCCATTAG
             JOZ241-C4       (1) ----------------------------------------
                                281                                    320
            AAO81020.1     (281) AACAATTCATGACAAAAGAATTAGTAGGATGCAGTAGTGA
JOZ241-C4 full-length      (281) AACAATTCATGACAAAAGAATTAGTAGGATGCAGTAGTGA
             JOZ241-C4       (1) ----------------------------------------
```

```
                                 321                                 360
            AAO81020.1    (321)  TTATGCAATTGCCGAAATGTTACTGAGTGGTATTACGAGC
JOZ241-C4 full-length     (321)  TTATGCAATTGCTGAAATGTTACTGAGTGGTATTACGAGC
            JOZ241-C4       (1)  ------------------------------------
                                 361                                 400
            AAO81020.1    (361)  TTTTGTGATATGTATTATTTTGAAGATGAAATTGCTAAAA
JOZ241-C4 full-length     (361)  TTTTGTGATATGTATTATTTTGAAGATGAAATTGCTAAAA
            JOZ241-C4       (1)  ------------------------------------
                                 401                                 440
            AAO81020.1    (401)  GTTGTGAAAAAATGAGTGTTCGTGCTTTGCTCGGAGAGAC
JOZ241-C4 full-length     (401)  GTTGTAAAAAAATGAGTGTTCGTGCTTTGCTCGGAGAGAC
            JOZ241-C4       (1)  ------------------------------------
                                 441                                 480
            AAO81020.1    (441)  GATCATTGATATGCCCACTTGTGATAGTCCGGAGCCTTCA
JOZ241-C4 full-length     (441)  GATCATTGATATGCCCACTTGTGATAGTCCGGAGCCTTCA
            JOZ241-C4       (1)  ------------------------------------
                                 481                                 520
            AAO81020.1    (481)  GGCGGTCTTTTTTACGCGGAAACCTTTATTCGCAAGTGGC
JOZ241-C4 full-length     (481)  GGCGGTCTTTTTTACGCGGAAACCTTTATTCGCAAGTGGC
            JOZ241-C4       (1)  ------------------------------------
                                 521                                 560
            AAO81020.1    (521)  AAGGCCATCCGTTGATTACGCCTATGCTTGCGCCACATGC
JOZ241-C4 full-length     (521)  AAGGCCATCCGTTGATTACGCCTATACTTGCGCCACATGC
            JOZ241-C4       (1)  ------------------------------------
                                 561                                 600
            AAO81020.1    (561)  ACCGAATACCAATTCACCAGAAGTGTTGGCGAAAATTATT
JOZ241-C4 full-length     (561)  ACCGAATACCAATTCACCAGAAGTGTTGGCGAAAATTATT
            JOZ241-C4       (1)  ------------------------------------
                                 601                                 640
            AAO81020.1    (601)  GAACTTAGTCGGCAATACCAAGTTCCTGTGACCATGCACG
JOZ241-C4 full-length     (601)  GAACTTAGTCGGCGATACCAAGTTCCTGTGACCATGCACG
            JOZ241-C4       (1)  ------------------------------------
                                 641                                 680
            AAO81020.1    (641)  TTGCTGAAATGACTTATGAAATGGCTGAGTTTGAAAAAGC
JOZ241-C4 full-length     (641)  TTGCTGAAATGACTTATGAAATGGCTGAGTTTGAAAAAGC
            JOZ241-C4       (1)  ------------------------------------
                                 681                                 720
            AAO81020.1    (681)  CTATCAAAAAACACCAATTGCTTTCTTAGAAGAACTGGGT
JOZ241-C4 full-length     (681)  CTATCAAAAAACACCAATTGCTTTCTTAGAAGAACTGGGT
            JOZ241-C4       (1)  ------------------------------------
                                 721                                 760
            AAO81020.1    (721)  TATTTGAGCGAGCCGTTTATTTTAGCGCATTGTATTTTGG
JOZ241-C4 full-length     (721)  TATTTGAGCGAGCCGTTTATTTTAGCTCATTGTATTTTGG
            JOZ241-C4       (1)  --------------------------------GTATTTTGG
                                 761                                 800
            AAO81020.1    (761)  CAACAGATGAAGATCTTGCGAGTTTAGCTGCTACTAATGG
JOZ241-C4 full-length     (761)  CAACAGATGAAGATCTTGCGAGTTTAGCTGCTACTAATGG
            JOZ241-C4      (10)  CAACAGATGAAGATCTTGCGAGTTTAGCTGCTACTAATGG
                                 801                                 840
            AAO81020.1    (801)  AAAAGCGCGTGTCGCTCATTGTATCGGTGCGAATACTAAA
JOZ241-C4 full-length     (801)  AAAAGCGCGTGTCGCTCATTGTATCGGTGCGAATACTAAA
            JOZ241-C4      (50)  AAAAGCGCGTGTCGCTCATTGTATCGGTGCGAATACTAAA
                                 841                                 880
            AAO81020.1    (841)  TCAGCCAAAGGCGTAGCGCCGATTAAGCAAATGCTTGATC
JOZ241-C4 full-length     (841)  TCAGCCAAAGGCGTAGCGCCGATTAAGCAAATGCTTGATC
            JOZ241-C4      (90)  TCAGCCAAAGGCGTAGCGCCGATTAAGCAAATGCTTGATC
```

```
                                 881                                       920
            AAO81020.1    (881) AAGGGATTATTGTCGGTTTAGGCACGGATGGACCTAGTAG
JOZ241-C4 full-length     (881) AAGGGATTATTGTCGGTTTAGGAACGGATGGACCTAGTAG
             JOZ241-C4    (130) AAGGGATTATTGTCGGTTTAGGAACGGATGGACCTAGTAG
                                 921                                       960
            AAO81020.1    (921) TGGGAATACATTAGATTTATTCACCCAAATGCGCATGGTT
JOZ241-C4 full-length     (921) TGGGAATACATTAGATTTATTCACCCAAATGCGCATGGTT
             JOZ241-C4    (170) TGGGAATACATTAGATTTATTCACCCAAATGCGCATGGTT
                                 961                                      1000
            AAO81020.1    (961) GCGAATTTTCATAAGACAGCACACCAAGATCGCTCCTTGT
JOZ241-C4 full-length     (961) GCGAATTTTCATAAGACAGCACACCAAGATCGCTCCTTGT
             JOZ241-C4    (210) GCGAATTTTCATAAGACAGCACACCAAGATCGCTCCTTGT
                                1001                                      1040
            AAO81020.1   (1001) TTCCTGCTAAAGAAATTGTTTATCTGGCAACGATGGGGGG
JOZ241-C4 full-length    (1001) TTCCTGCTAAAGAAATTGTTTATCTGGCAACGATGGGGGG
             JOZ241-C4    (250) TTCCTGCTAAAGAAATTGTTTATCTGGCAACGATGGGGGG
                                1041                                      1080
            AAO81020.1   (1041) CGCTAAAACGTTAGGCTTGGCGGAGCAAGTCGGCTCATTG
JOZ241-C4 full-length    (1041) CGCTAAAACGTTAGGCTTGGCGGAGCAAGTCGGTTCATTG
             JOZ241-C4    (290) CGCTAAAACGTTAGGCTTGGCGGAGCAAGTCGGTTCATTG
                                1081                                      1120
            AAO81020.1   (1081) GAAGTGGACAAAAAAGCGGATATAACATTAATTGAAACGC
JOZ241-C4 full-length    (1081) GAAGTGGGCAAAAAAGCGGATATAACATTAATTGAAACGC
             JOZ241-C4    (330) GAAGTGGGCAAAAAAGCGGA--------------------
                                1121                                      1160
            AAO81020.1   (1121) AATCAGTTAATATGTTTCCGATTTTTGATGCCTATTCAGC
JOZ241-C4 full-length    (1121) AATCAGTTAATATGTTTCCGATTTTTGATGCCTATTCAGC
             JOZ241-C4    (350) ----------------------------------------
                                1161                                      1200
            AAO81020.1   (1161) GTTGGTTTATTCAGCAAATGCTAGCAATGTTGAAGCCGTT
JOZ241-C4 full-length    (1161) ATTGGTTTATTCAGCAAATGCTAGCAATGTTGAAGCCGTT
             JOZ241-C4    (350) ----------------------------------------
                                1201                                      1240
            AAO81020.1   (1201) TGGGTGAACGGTCAACAGTTAGTTGCCAATAAAGAATTAC
JOZ241-C4 full-length    (1201) TGGGTGAACGGTCAACAGTTAGTTGCCAATAAAGAATTAC
             JOZ241-C4    (350) ----------------------------------------
                                1241                                      1280
            AAO81020.1   (1241) AACAAGCTAATCTCAAAGAAATCAAGGAAAAATTATATCA
JOZ241-C4 full-length    (1241) AACAAGCTAATCTCAAAGAAATCAAGGAAAAATTATATCA
             JOZ241-C4    (350) ----------------------------------------
                                1281                                      1320
            AAO81020.1   (1281) GGCCATGAATACGTTTGTGAAAGAAGCTAAAAAAAGAGCT
JOZ241-C4 full-length    (1281) GGCCATGAATACGTTTGTGAAAGAAGCTAAAAAAAGAGCT
             JOZ241-C4    (350) ----------------------------------------
                                1321                                      1360
            AAO81020.1   (1321) GCTCTC------------------------------------
JOZ241-C4 full-length    (1321) GCTCTCGCGGCCGCACTCGAGCACCACCACCACCACCACT
             JOZ241-C4    (350) ----------------------------------------
                                1361
            AAO81020.1   (1327) --
JOZ241-C4 full-length    (1361) GA
             JOZ241-C4    (350) --
```

11.4.3.2 Protein alignment



11.4.4 JOZ242-C9

11.4.4.1 DNA alignment

```
                                1                                        40
            ALE13104.1     (1)  ATGTGCAGGAGATTCGCGCTCGATTTGGATTGGGATGCCG
JOZ242-C9 full-length      (1)  ATGTGCAGGAGATTCGCGCTCGACTTGGATTGGGATGCCG
             JOZ242-C9     (1)  ----------------------------ATTGGGATGCCG
                               41                                       80
            ALE13104.1    (41)  TGGCCTCACAGTTCGCGGTAGACGAGGATGACGTGCGCGT
JOZ242-C9 full-length     (41)  TGGCCTCACAGTTCGCGGTAGACGAGGATGACGTGCGCGT
             JOZ242-C9    (13)  TGGCCTCACAGTTCGCGGTAGACGAGGATGACGTGCGCGT
                               81                                      120
            ALE13104.1    (81)  TGACACTCTGCCACAGCCTTCCTACAACATCGCACCCACG
JOZ242-C9 full-length     (81)  TGACACTCTGCCACAGCCTTCCTACAACATCGCACCCACG
             JOZ242-C9    (53)  TGACACTCTGCCACAGCCTTCCTACAACATCGCACCCACG
                              121                                      160
            ALE13104.1   (121)  CAGAACATTGGCGTGGTGGCACAGGGCAAGGATGGCCGCC
JOZ242-C9 full-length    (121)  CAGAACATTGGCGTGGTGGCACAGGGCAAGGATGGCCGCC
             JOZ242-C9    (93)  CAGAACATTGGCGTGGTGGCACAGGGCAAGGATGGCCGCC
                              161                                      200
            ALE13104.1   (161)  GTCATCTGACCGGTGCCTATTGGTCGCTGGTGCCGCGATG
JOZ242-C9 full-length    (161)  GTCATCTGACCGGTGCCTATTGGTCGCTGGTGCCGCGATG
             JOZ242-C9   (133)  GTCATCTGACCGGTGCCTATTGGTCGCTGGTGCCGCGATG
                              201                                      240
            ALE13104.1   (201)  GAGCGCCAGCAAAGTGCTGAGCTACCCCACATATAATGCG
JOZ242-C9 full-length    (201)  GAGCGCCAGCAAAGTGCTGAGCTACCCCACATATAATGCG
             JOZ242-C9   (173)  GAGCGCCAGCAAAGTGCTGAGCTACCCCACATATAATGCG
                              241                                      280
            ALE13104.1   (241)  CGTGTGGAGTCTGCGCATGTCAAGCCGGCGTTCGCCGAAT
JOZ242-C9 full-length    (241)  CGTGTGGAGTCTGCGCATGTCAAGCCGGCGTTCGCCGAAT
             JOZ242-C9   (213)  CGTGTGGAGTCTGCGCATGTCAAGCCGGCGTTCGCCGAAT
                              281                                      320
            ALE13104.1   (281)  CCACAAAATCCATGCGAGCCATCATCCCCGCTTCCGGCTA
JOZ242-C9 full-length    (281)  CCACAAAATCCATGCGAGCCATCATCCCCGCTTCCGGCTA
             JOZ242-C9   (253)  CCACAAAATCCATGCGAGCCATCATCCCCGCTTCCGGCTA
                              321                                      360
            ALE13104.1   (321)  CTACGAATGGAAAGGTCGCAGGCCGTTCTATTTTCACGCA
JOZ242-C9 full-length    (321)  CTACGAATGGAAAGGTCGCAGGCCGTTCTATTTTCACGCA
             JOZ242-C9   (293)  CTACGAATGGAAAGGTCGCAGGCCGTTCTATTTTCA----
                              361                                      400
            ALE13104.1   (361)  CCACATGATGAACTATTGTCTTTGGCCGGCCTGTACTCAT
JOZ242-C9 full-length    (361)  CCACATGATGAACTATTGTCTTTGGCCGGCCTGTACTCAT
             JOZ242-C9   (329)  ----------------------------------------
                              401                                      440
            ALE13104.1   (401)  GGTGGAGGCCGTCTCCCGCATCGCCATGGCAGCTGACCGC
JOZ242-C9 full-length    (401)  GGTGGAGGCCGTCTCCCGCATCGCCATGGCAGCTGACCGC
             JOZ242-C9   (329)  ----------------------------------------
                              441                                      480
            ALE13104.1   (441)  AACCATCATCACCTGCCCTGCCGCGGACGAGTTCGTCAAG
JOZ242-C9 full-length    (441)  AACCATCATCACCTGCCCTGCCGCGGACGAGTTCGTCAAG
             JOZ242-C9   (329)  ----------------------------------------
                              481                                      520
            ALE13104.1   (481)  GTGCATGATCGCATGCCCCTGCTGGCTCCTCGGAATATGG
JOZ242-C9 full-length    (481)  GTGCATGATCGCATGCCCCTGCTGGCTCCTCGGAACATGG
             JOZ242-C9   (329)  ----------------------------------------
```

```
                                521                                         560
           ALE13104.1   (521) TTGCCTCATGGCTGGACCGCTCCGTTGACGGGGCCGCACT
JOZ242-C9 full-length   (521) TTGCCTCATGGCTGGACCGCTCCGTTGACGGAGCCGCACT
            JOZ242-C9   (329) ----------------------------------------
                                561                                         600
           ALE13104.1   (561) GTTAGATTCCATGCGGGAAGCGGGTACCATGCTTTCCCGG
JOZ242-C9 full-length   (561) GTTAGATTCCATGCGGGAAGCGGGTACCATGCTTTCCCGG
            JOZ242-C9   (329) ----------------------------------------
                                601                                         640
           ALE13104.1   (601) CGTCTGCAATTTCACGAAGTCGCGCCACTGAACAGCGACG
JOZ242-C9 full-length   (601) CATCTGCAATTTCACGAAGTCGCGCCACTGAACAGCGACG
            JOZ242-C9   (329) ----------------------------------------
                                641                                         680
           ALE13104.1   (641) GCAAACGTCTTATTCAGCCATTGGATTCCACCGAACCCAT
JOZ242-C9 full-length   (641) GCAAACGTCTTATTCAGCCATTGGATTCCACCGAACCCAT
            JOZ242-C9   (329) ----------------------------------------
                                681                                         720
           ALE13104.1   (681) GCGTCTGTTT------------------------------
JOZ242-C9 full-length   (681) GCGTCTGTTTGCGGCCGCACTCGAGCACCACCACCACCAC
            JOZ242-C9   (329) ----------------------------------------
                                721
           ALE13104.1   (691) ------
JOZ242-C9 full-length   (721) CACTGA
            JOZ242-C9   (329) ------
```

11.4.4.2 Protein alignment

```
                                  1                                          40
           ALE13104.1     (1) MCRRFALDLDWDAVASQFAVDEDDVRVDTLPQPSYNIAPT
JOZ242-C9 full-length     (1) MCRRFALDLDWDAVASQFAVDEDDVRVDTLPQPSYNIAPT
            JOZ242-C4     (1) ----------WDAVASQFAVDEDDVRVDTLPQPSYNIAPT
                                 41                                          80
           ALE13104.1    (41) QNIGVVAQGKDGRRHLTGAYWSLVPRWSASKVLSYPTYNA
JOZ242-C9 full-length    (41) QNIGVVAQGKDGRRHLTGAYWSLVPRWSASKVLSYPTYNA
            JOZ242-C4    (31) QNIGVVAQGKDGRRHLTGAYWSLVPRWSASKVLSYPTYNA
                                 81                                         120
           ALE13104.1    (81) RVESAHVKPAFAESTKSMRAIIPASGYYEWKGRRPFYFHA
JOZ242-C9 full-length    (81) RVESAHVKPAFAESTKSMRAIIPASGYYEWKGRRPFYFHA
            JOZ242-C4    (71) RVESAHVKPAFAESTKSMRAIIPASGYYEWKGRRPFYF--
                                121                                         160
           ALE13104.1   (121) PHDELLSLAGLYSWWRPSPASPWQLTATIITCPAADEFVK
JOZ242-C9 full-length   (121) PHDELLSLAGLYSWWRPSPASPWQLTATIITCPAADEFVK
            JOZ242-C4   (109) ----------------------------------------
                                161                                         200
           ALE13104.1   (161) VHDRMPLLAPRNMVASWLDRSVDGAALLDSMREAGTMLSR
JOZ242-C9 full-length   (161) VHDRMPLLAPRNMVASWLDRSVDGAALLDSMREAGTMLSR
            JOZ242-C4   (109) ----------------------------------------
                                201                                         240
           ALE13104.1   (201) RLQFHEVAPLNSDGKRLIQPLDSTEPMRLF----------
JOZ242-C9 full-length   (201) HLQFHEVAPLNSDGKRLIQPLDSTEPMRLFAAALEHHHHH
            JOZ242-C4   (109) ----------------------------------------
                                241
           ALE13104.1   (231) -
JOZ242-C9 full-length   (241) H
            JOZ242-C4   (109) -
```